Rationality and Irrationality

Proceedings of
The 23rd International Wittgenstein-Symposium

13th to 19th August 2000
Kirchberg am Wechsel (Austria)

Editors
BERIT BROGAARD
BARRY SMITH

Vienna 2001
öbv&hpt
Verlagsgesellschaft

Rationalität und Irrationalität

Akten des
23. Internationalen Wittgenstein-Symposiums

13. bis 19. August 2000
Kirchberg am Wechsel (Österreich)

Herausgeber
BERIT BROGAARD
BARRY SMITH

Wien 2001
öbv&hpt
Verlagsgesellschaft

Schriftenreihe der Wittgenstein-Gesellschaft

Volume 29

Wir danken
dem Bundesministerium für Bildung, Wissenschaft und Kultur in Wien
und dem Kulturamt der Landesregierung Niederösterreich in St. Pölten
für die Förderung dieses Werkes

Die Deutsche Bibliothek – CIP-Einheitsaufnahme

Ein Titeldatensatz für diese Publikation ist bei Der
Deutschen Bibliothek erhältlich

ISBN 3-209-03648-9

Umschlaggestaltung: Pia Moest
Layout: Thomas Binder
Gesamtherstellung: Novographic, Wien

Table of Contents
Inhaltsverzeichnis

Vorwort
Preface 9

Black Swans
The formative influences in Australian philosophy
DAVID M. ARMSTRONG 11

Rationality and Reasonableness
MICHAEL BEANEY 18

Thomas Hobbes: The Rationalization of Religion
ANAT BILETZKI 25

Elusive Reference
BERIT BROGAARD 36

Sind alle Religionen gleich rational?
ANDRZEJ BRONK 46

Posthumanism: Engineering in the Place of Ethics
STEPHEN CLARK 62

A Profession of Stupidity
RONALD DE SOUSA 77

Rationality and Irrationality in Scientific Language
LUIS FLORES H. 94

Oxford Philosophy: A Case Study in Cognitive Epidemiology
LYND FORGUSON 101

The Rationality of Epistemology and the Rationality of Ontology
ANDREW FRANK 110

Reason and Necessity
NEWTON GARVER 120

Causal Domains and Emergent Rationality
IVAN M. HAVEL 129

Reasons, Truthmakers and Truth Grounds
HERBERT HOCHBERG 152

The Two-Envelope Paradox
and the Foundations of Rational Decision Theory
TERRY HORGAN 172

The Rationality of Reasoning: Commitment and Coherence
JOHN KEARNS 192

Welche Stufe der Rationalität ist in Recht und Ethik
erreichbar und wünschenswert?
EDGAR MORSCHER 199

The Invention of Western Reason
PHILIPPE NEMO 224

The Picture Theory of Reason
J. C. NYÍRI 242

The Irrationality of Religion
A Plea for Atheism
HERMAN PHILIPSE 267

Philosophy in Finland – Analytic and Post-Analytic
SAMI PIHLSTRÖM 273

Why It's Irrational to Believe in Consistency
GRAHAM PRIEST 284

Rationalität und der Glaube der „religiös Eingeweihten"
EDMUND RUNGGALDIER 294

Kinds of Rationality and their Role in Evolution
GERHARD SCHURZ 301

The Classical Model of Rationality and Its Weaknesses
JOHN R. SEARLE 311

Wann ist die Vernunft praktisch und wann ist Normativität moralisch?
ULRICH STEINVORTH 325

Schemata, Abstraction, and Biology
Man as the Abstract Animal rather than the Symbolic Species?
FREDERIK STJERNFELT 341

Privileged Rationality
AVRUM STROLL 362

Multiplicity of Mental Spaces
BARBARA TVERSKY 370

Kurz, knapp, konsistent?
Schwierigkeiten mit einem regulativen Ideal
MAX URCHS 379

Rationalism and Irrationalism: The Case of Poland
JAN WOLEŃSKI 390

Autorenverzeichnis
List of authors 407

Preface

The 23rd International Wittgenstein Conference, which took place in Kirchberg am Wechsel, Lower Austria in the week of 13–19 August 2000, was devoted to the topic of *Rationality and Irrationality*. Almost 100 talks were presented at the conference, of which the present volume represents primarily the invited papers. The submitted papers have appeared already in:

> Berit Brogaard (ed.), *Rationality and Irrationality* (Contributions of the Austrian Ludwig Wittgenstein Society, 8), 2 volumes, Kirchberg am Wechsel: Österreichische Ludwig Wittgenstein Gesellschaft, 2000.

In organizing the conference we sought to explore the role played by rationality in different areas of contemporary philosophy, and to explore what exactly rationality is.

Among the topics treated were: truth, psychologism, science, the nature of rational discourse, practical reason, contextualism, vagueness, types of rationality, the rationality of religious belief, and Wittgenstein. Questions addressed included: Is rationality tied to special sorts of contexts? Is rationality tied to language? Is scientific rationality the only kind of rationality? Is there something like a *Western* rationality? and: Could we genetically engineer human beings to be less wicked? The opening lecture at the conference was delivered by John Searle, who set the scene for the rest of the week by presenting arguments, taken from his new book *Rationality in Action*, against what he sees as the still dominant 'classical' view of rationality. The latter sees rationality as a matter of employing logical reasoning in determining the best means to achieve a given end. To be rational, on this view, is to have a certain consistent set of desires and to obey logical rules in determining how to act. Rationality thus relates to the means for achieving ends which have been somehow pre-determined. On Searle's own view, in contrast, rationality can and must apply to the determination of ends just as much to the selection of means. Rationality consists, according to Searle, not in obeying rules, but rather in the exercise of the free will of the rational self – a theme which served as one important undercurrent through the conference as a whole.

We had hoped that the broader title, *Rationality and Irrationality*, would encourage analytically minded philosophers to reflect on why it is that nonsense, stupidity, and sheer bad philosophy should have proved so consistently popular not only among the wider intellectual public but also in some circles of philosophy itself. We had hoped, in other words, that the defenders of rationality would take up the cudgels against the dark forces of irrationality which are in our midst.

But such was not to be. Rather, of the far more than 200 submissions which we received, a significant number consisted in more or less ingenious defenses of *irrationality*. Not all of the latter were rejected by the committee charged with refereeing submitted papers, but many of them (on topics like: woman and sex in post-modernist theater) were.

In his *The Four Phases of Philosophy* Franz Brentano presents a picture of the history of philosophy as an ever-repeating cycle of ups and downs. Up-phases turn on the domi-

nance of an empirical, scientific and logical orientation; down-phases on the dominance of irrationality and obfuscation and of politico-ideological distortion. The late twentieth-century phase in the history of philosophy, with its Rortys and its Derridas, must – at least when judged on the basis of its intellectual fashions – be counted as a miserable down-phase. We hope, however, that the serious work on behalf of rationality presented in this volume will serve as a reminder that there is still, behind the sham, considerable rational substance left to contemporary philosophy – and that fashions do after all wane.

Berit Brogaard and Barry Smith,
Rochester and Koblenz, November 2000

Black Swans
The formative influences in Australian philosophy

DAVID M. ARMSTRONG

I will start with an extract from the magazine *Why?*, edited, I think, by Anthony Kenny, published in Oxford in 1958. Mr L. Sturch there maintains that it is a fundamental error to think that:

> ... the question "Is there any reason for saying that in Australia the winter is in the summer?" has the same logic as "Is there any reason for saying that in France frogs are esteemed as a source of food?" It is a mistake to think that the name "Australia" has the same logical grammar as "France", "Switzerland", "Siberia", "Rutlandshire", or "North Dakota". It is no more like such names than "Utopia", "Erewhon", or "Ruritania" are. It is not sense to say "In Ruritania the population is increasing" *unless* you are playing a language-game in which it is stipulated that Ruritania is "a real place" (to use the material mode). Now it is clear that "Australia" is *not* a real place; or better, that the word "Australia" is not a name. The words "in Australia" are used simply to signify that the contradictory of what is stated to be the case "in Australia" is in fact the case. Thus we say "In Australia there are mammals that lay eggs" (meaning that there are none in reality); "In Australia there are black swans" (meaning that all real swans are some other colour); "In Australia people who stand upright have their heads pointing downwards" (meaning that this is self-contradictory).

Against the ingenious L. Sturch, with his theory that the phrase 'In Australia' is a negation-operator applied to sentences, I maintain that Australia really exists, and, perhaps a little suprisingly, that it contains quite a large number of philosophers. Indeed, these, like some other Australians, have been quite noisy.

There is a good book by Selwyn Grave *A History of Philosophy in Australia* which came out in 1984, and which takes the story up to 1980. It not only deals with the thought and teaching of the philosophers of Australia, which is its main concern, but also gives some account of the academic-cum-political struggles that enlivened the Australian philosophical scene. There is also in preparation a book by Jim Franklin to be called *Corrupting the Youth*, to be published by Melbourne University Press, one chapter of which has already been published (Franklin 1999). This, which goes over much of the same ground, often in more polemical fashion, should be of great interest.

What I will do in this little talk, is to give you some feeling for what I take to be the three great formative events in Australian philosophy: John Anderson in Sydney, George Paul and later Wittgensteinians in Melbourne – in particular Douglas Gasking – and Ullin Place, Jack Smart, and Charlie Martin in Adelaide.

Although the continent of Australia (we prefer being the smallest continent to being

B. Brogaard, B. Smith (Hg.), *Rationality and Irrationality / Rationalität und Irrationalität*, S. 11–17.

the largest island) is very old, something that visitors of any sensitivity quickly realize, nevertheless politically and institutionally it is very young. The nation, indeed, only came into existence on the first day of 1901. Before that there were only Crown Colonies, the oldest, New South Wales, dating to 1788. We are, in every sense, a transported civilization. (An English joke. The Englishman arrives at Sydney airport. The immigration officer says to him: 'Have you a criminal record?' 'I didn't think that was still necessary.') So we long depended for our first universities on those who came from the north, and in particular from Britain. Sydney University actually began in 1851, but adopted a latin motto that may be translated as 'The same mind under different skies'.

A Chair of Philosophy was established at Sydney in the 1880s, but things really started to move with the arrival of John Anderson in 1927. He had been born in 1893, did his undergraduate work at Glasgow University, going on to become a lecturer at Edinburgh in Norman Kemp Smith's department. I have to declare an interest here, because I received my intellectual formation from Anderson, but I think that he was by far the most remarkable figure that Australian philosophy has seen.

His thinking was systematic, and extraordinarily wide ranging, especially given our contemporary philosophical perspective. He had worked-out views in metaphysics, in epistemology, in the philosophy of mind, in morals, politics and social theory, in aesthetics and literature, and pretty much anything that might engage the intellectuals of his time. Marx, Freud and James Joyce were of particular interest to him. In provincial Sydney of those days he could give you an education that was far wider than Western philosophy narrowly conceived.

William James distinguished between the tough and the tender-minded philosophers. There seemed to be no atom of tender-mindedness in Anderson's thinking. There was, he held, only a 'single level of being', a world of continuously interacting situations in space and time. Minds, knowledge, morality, education, society, were no more than empirical realities, spatiotemporal realities that the inquiring mind might investigate, seeking to strip away the illusions that hung about them. Social life was not some unified affair, but a continuous interaction of different social movements with different, often irreconcilable, ways of life.[1] The life of inquiry, which he championed, was no more than a particular way of life.

When Anderson arrived in Australia he was an outspoken political radical, happy to support Leninism in the shape of the Australian Communist party (he even acted as an adviser to the Central Committee). It was a cause of scandal and concern in the university and the town. But he never subordinated his own political and philosophical judgement to communism. And remarkably early among intellectuals of the left all over the world, he became a critic of Stalinism. In the next few years he turned towards the Trotskyite position, but by the end of the thirties he had become a bitter critic of revolutionary socialism. That became for him one of the great illusions that the inquirer had to see through. He

1. Anderson wrote mostly for the *Australasian Journal of Psychology and Philosophy* (a journal that became the *Australasian Journal of Philosophy*). His papers were collected in a book: *Studies in Empirical Philosophy* in 1962, the year of his death. But those who were not his students find him hard to read. A.J. Baker has written two very clear and accurate books on Anderson's social philosophy (1979) and his systematic philosophy (1986).

never became an orthodox conservative. Freedom in general, and academic freedom in particular, would always, he thought, lead 'a perilous and fighting life'. But it is 'no accident' as the Stalinists used to say, that among the very large number of intellectuals of Sydney that he taught and influenced, quite a substantial proportion turned to some variety of liberal-conservative political position.

One bad thing, as I think, about Anderson is that, although he preached critical inquiry, he was very intolerant of it when it was directed against his own views, especially in his own department and from his own students. (In this he resembled another outspoken upholder of critical inquiry: Karl Popper. I sometimes wonder whether there is a law of nature here.) Whether he admitted it to himself or not, what Anderson really wanted people to get was a good grip of, and acceptance of, his own position. The Andersonians, as those who bought the whole, or nearly the whole, package were called, were typical *disciples*. But if you had the strength of mind, or character, or just plain cussedness, to learn from Anderson without becoming subject to him, you could get a wonderful education from him. John Passmore, John Mackie, David Stove, Eugene Kamenka and me, although we went different ways, could all attest to this.

One good, indeed excellent, thing about Anderson was that he saw philosophy historically. I do not mean that he saw it in a scholarly way, he was no great scholar, but rather he saw it as a great argument that had been going on since Thales. I think that he thought that in the long procession of philosophers up to present times there were only two who had really achieved a true view of things: himself and, in a more intuitive way, Heraclitus. Fragment 20, in John Burnet's translation (1925), I suppose sums up in brief Heraclitus' metaphysics. It would do for Anderson as well:

> This world, which is the same for all, no one of gods or men has made; but it was ever, is now, and ever shall be an ever-living Fire, with measures of it kindling, and measures going out.

The measures may be read as Anderson's 'ways of working' of things, their immanent laws.

For many, it was Anderson's social views that were of the greatest interest. For me it was the metaphysics. The position that there is nothing more to being than the spatiotemporal system is hardly an outlandish view. For an irreligous person, I suppose, it is no more than commonsense. Just brass tacks, as Ryle said in an article on Anderson (1950). Still and all, it's a great way to start. For, after all, Anderson being a philosopher, necessarily faced two ways: not only rejecting any sort of deity, but also the extraordinary entities postulated by so many philosophers, from the time of Plato onwards, at least, right up to the present day. Furthermore, being a traditional philosopher, Anderson was not going to leave spacetime just to the scientists. He wanted to put forward a particular view of its most abstract structure.

He argued that reality had a *propositional* structure. By this he meant nothing idealistic and nothing linguistic. Perhaps his idea is best understood today by saying that for him the world is a world of facts rather than a world of things. He was here aligning himself with the logical atomism of Russell and the *Tractatus* Wittgenstein. I have said 'align' but Anderson never really aligned himself intellectually with anybody. He rejected atom-

ism in favour of a doctrine of the infinite complexity of things, and he never really came to grips with Wittgenstein, early or late, though he was scathing about the 'linguistic turn'. The resemblances between his ontological views and the doctrines of the *Tractatus* were pointed out by Douglas Gasking in a paper published in 1949. My own 'states of affairs' are directly descended from Anderson's propositional view of reality.

Anderson never accepted the new Russellian logic either, and sought, implausibly, to exhibit all propositions as falling under the Aristotelian 'four forms' of subject-predicate propositions. This led to trouble with relations, whose reality Anderson was most keen to uphold, but they had to be smuggled into the four forms as relational properties. Anderson did go on to an interesting theory of categories, taking his lead from Samuel Alexander (who, by mere coincidence, was born in Australia although he spent his student and academic life in England). Alexander developed a realist treatment of space, time and the categories of being, putting an anti-subjectivist transformation on Kant. Anderson heard Alexander's Gifford lectures, published as *Space, Time and Deity* (1920). Anderson had no use for Alexander's emergent and non-transcendent deity, but he was taken by the idea of the categories. His own system included thirteen categories, which included Identity, Difference, Existence, Quality, Relation, Number, Quantity, Intensity and Causality, all linked to the form of the proposition. Anderson's dictated lectures on Alexander exist, but have never been published. They might gain more attention now that the idea of an empirical metaphysics has returned, in a modest way, to the philosophical agenda.

I could go on about Anderson for a long time, but let me now turn to Wittgensteinian Melbourne. The seminal figure here was the Englishman George Paul. Paul was at Melbourne University during the years of the second world war, returning to England after that. He brought the philosophy of Wittgenstein, whose pupil he had been, to Melbourne University. Though the period was relatively short his influence during that time was immense, and not merely within the philosophy department. I have never read or heard a detailed account of those years, but a friend of mine who became an anthropologist told me that it was as if the philosopher-king had arrived in Melbourne. His influence on the whole Faculty of Arts was, it appears, immense. To the historic rivalry between Sydney and Melbourne, a rivalry having a great number of dimensions, a rivalry to which we owe the mid-point location and still artificial nature of our capital, Canberra, was added a new and not unimportant dimension: Andersonianism versus Wittgensteinianism.

Paul's influence in Melbourne lived on, for instance in A.C. 'Camo' Jackson, the father of Frank Jackson. But the Wittgenstein influence was further secured by the arrival in 1946 of another Englishman, and pupil of the master, Douglas Gasking, who spent the rest of his life in Melbourne. Gasking perhaps succumbed a little to what one of our national poets described as the Australian 'dream of ease'. But in all his lecturing and in his relatively small number of publications he sought, and achieved, total clarity. The most important of these have at last been collected in a book *Language, Logic and Causation* (1996). His thought did develop over the years. He became sympathetic to 'Australian materialism', and Quine and Davidson were also influences.

After the end of the war began an Age which still continues, the Age of the Conference. In Australian philosophy this meant, effectively, the annual coming together of, or rather the clashing of, two philosophically self-confident groups, one from Melbourne, one from Sydney, who did not find it easy to understand each other. There was a good

deal of intransigence, particularly on the part of Sydney, I think. Anderson had evolved a style of giving papers that was (as far as I know) all his own. When the discussion began, he took careful notes, but did not speak in reply until everybody else had finished. Then he gave a speech in reply, in the course of which he took note of, and in general criticized, what had been said. It enabled him to re-emphasize his main themes, which as we all know can get lost in ordinary back-and-forth question and answer. But it didn't help much to get the detail of arguments straight, which is also very important for philosophers. Gasking, in particular, was much more eirenic, genuinely seeking to understand and even to find common ground. (I have already mentioned his paper comparing and contrasting Anderson's position and that of the *Tractatus*.) And some of us genuinely wanted to find out what was going on in Melbourne, and read with great interest a typescript of Wittgenstein's *Blue Book*, which circulated rather clandestinely in Sydney.

But even after all of this has been said, the intellectual temper of the two schools was very far apart. The idea that philosophy was a sort of muddle which needed clearing up – the fly shown the way out of the fly-bottle – and the linguistic turn that grew out of this, was deeply opposed to the traditional and classical conception of philosophy that, in an empiricist version, prevailed in Sydney. To a large extent it was a dialogue of the deaf. For the English-speaking philosophical world, the later Wittgenstein and linguistic philosophy were the fashion, Andersonianism could not have been less fashionable. Now that the fashion has passed, though it has left important marks, many Andersonian ideas can be put forward and discussed in a way that was hardly possible then.

In the early fifties, however, something new entered Australian philosophy. This was the arrival in 1950 of Jack Smart to the Chair of philosophy in Adelaide University, South Australia. Smart was very young, even younger than Anderson when the latter came out to Sydney. He had been a graduate student at Oxford, and was backed by Gilbert Ryle. Smart was a disciple of Ryle's and, in particular, accepted Ryle's philosophy of mind as set out in *The Concept of Mind*.

Smart's department was supposed to cover both philosophy and psychology, and he needed a psychologist. The person he appointed was Ullin Place, known to the anthologies as U.T. Place. He had also been at Oxford, taking the PPP course: Philosophy, Psychology and Physiology, a course taken by few, but those few rather select. He was recommended to Smart by Brian Farrell, the author of the rather remarkable paper 'Experience', remarkable, given its content, in being published in 1950. With its contention that experience is 'featureless', it anticipated Smart's doctrine of the topic neutrality of mental discourse, and, incidentally, introduced the question what it is like to be a bat. Place himself began experimental psychology at Adelaide, but the great contribution he made there was, of course, in the philosophy of mind.

Another appointment that Smart was able to make at that time was also very important. This was C.B. 'Charlie' Martin. An American, he had been a student of John Wisdom's at Cambridge. He rapidly became disillusioned with Wittgensteinian philosophy. He used to say that he wanted to know what there is, and his preoccupation with ontological issues was profoundly important in Adelaide and, later, in the wider Australian philosophical scene. It was Martin who introduced in Australia the concept of a truthmaker, that in the world, whatever it is, in virtue of which a true proposition is true. He first applied it to the counterfactuals about possible perceptions used by the Phenomenalists, and

the dispositional truths about behaviour which were so important in a Rylean philosophy of mind.

In Adelaide, Place started things going. He began, like Smart, from Rylean behaviourism. But the existence of inner mental processes seemed to him to be undeniable. At the same time, the arguments for a physicalist account of the world, including the mental, seemed very strong on scientific grounds. So he came to the Identity theory – the identification of mental processes with purely physical processes in the brain. It is important on the grounds of historical justice to realize something, something that Smart has constantly born witness to but the philosophical world has nevertheless often been confused about, that Place had to convert Smart from his Rylean view. A large part of the trouble arose, I think, because Place's 1956 paper was published in the *British Journal of Psychology*, which few philosophers read. Smart's famous paper did not appear until 1959, but it was in the *Philosophical Review*, and anybody who was anybody in analytic philosophy read it. Later both papers started appearing more or less side by side in anthologies, and perception of the direction of influence became a bit blurred.

Martin was never, I think, a Rylean, but he too sought to oppose Place's view, looking instead for some 'Double-Aspect' view. But the years in which these three powerful intelligences were arguing the matter back and forth in the Adelaide department, quite a brief time because Place went back to England in 1956 and was hardly ever in Australia again, constitute one of the heroic episodes in Australian philosophy, and I think one of the defining ones.

These, then, as I see it, are the three great formative influences on philosophy in Australia. Nowadays, of course, we are more like everybody else in the philosophical world. You can find all the fashionable approaches to the subject among Australian philosophers now, and all the fashionable topics, analytic and non-analytic. But perhaps a certain idiosyncratic intellectual temper remains, a certain flavour that still distinguishes us. If so, I think it springs from the influences of Anderson in Sydney; from George Paul, Douglas Gasking and the Wittgensteinian tradition in Melbourne; and from Place, Smart and Charlie Martin in Adelaide.

Literature

Alexander, Samuel, 1920: *Space, Time and Deity* (2 vols.), London: Macmillan.

Anderson, John,1962: *Studies in Empirical Philosophy*, Sydney: Angus & Robertson. (There is a useful preface by John Passmore.)

Baker, A.J. 1979: *Anderson's Social Philosophy*: The Social Thought and Political Life of Professor John Anderson, Sydney: Angus & Robertson Publishers.

—, 1986: *Australian Realism: The Systematic Philosophy of John Anderson*. Cambridge: Cambridge University Press.

Burnet, John 1945: *Early Greek Philosophy*, fourth edition, London: Adam & Charles Black.

Farrell, Brian 1950: "Experience". *Mind*, 59, 170–198.

Franklin, James 1999: "The Sydney Philosophy Disturbances". *Quadrant*, 43 (April), 16–21.

Gasking, D.A.T. 1949: "Anderson and the Tractatus Logico-Philosophicus", *Australasian Journal of Philosophy*, 27, 1–26.

Grave, S.A. 1984: *A History of Philosophy in Australia*, St. Lucia, Queensland (Australia): University of Queensland Press.

Oakley I.T, & O'Neill L.J. (eds.) 1996: *Language, Logic and Causation: Philosophical Writings of Douglas Gasking*, Melbourne: Melbourne University Press.

Ryle, Gilbert 1950: "Logic and Professor Anderson", *Australasian Journal of Philosophy*, 28, reprinted in Ryle's *Collected Papers*, Vol.1, London: Hutchinson, 1971, 236–248.

Rationality and Reasonableness*

MICHAEL BEANEY

This paper is concerned with the distinction between rationality and reasonableness, a distinction that is firmly entrenched in ordinary language but which seems to have suffered relative neglect in certain recent (analytic) philosophy.

1. The Rational and the Reasonable Person

Most people, I think, would prefer a boss who was 'reasonable' rather than 'rational'. The contrast is firmly rooted in everyday language. Someone who is 'reasonable' is someone who is prepared to listen, cooperate and compromise, and can make informed and balanced judgements. Someone who is 'rational' is someone who can think through the implications of something, clearly and consistently, from principles already laid down, and come to the best decision possible on the evidence at hand, but who does not necessarily have the 'imagination' or 'emotional intelligence' to appreciate the wider picture or take into account practicalities, future possibilities or natural human reactions (which might supply further relevant evidence). The 'rational' is often contrasted to the 'emotional', whilst the 'reasonable' is contrasted to the 'insensitive', 'inhuman' or 'unrealistic', contrasts reflected in such phrases as "That may be rational, but it is very unreasonable".

But whilst the contrast is often drawn, is there any necessary conflict between rationality and reasonableness, once we ignore the negative connotations of 'rational'? Shouldn't the ideal boss be both rational and reasonable? If 'rational' is more a logical or analytic concept and 'reasonable' more an ethical or hermeneutic concept, is there any intrinsic incompatibility?

2. Gilbert Harman on Rationality: A Reasonable Discussion?

In his recent discussion of rationality, Gilbert Harman (1999, pp. 10, 43) gives the following example:

* This paper was written whilst a Research Fellow at the Institut für Philosophie of the University of Erlangen-Nürnberg, funded by the Alexander von Humboldt-Stiftung. I am grateful to both institutions for their generous support. The paper was read at both the Colloquium Logico-Philosophicum at Erlangen in April 2000 (in a German version) and the 23rd International Wittgenstein Symposium in Kirchberg am Wechsel in August 2000. I would like to thank participants at both for comments, and the present version incorporates a number of minor revisions and six new notes.

B. Brogaard, B. Smith (Hg.), *Rationality and Irrationality / Rationalität und Irrationalität*, S. 18–24.

> *Refusing a reasonable proposal*
> Three students, Sally, Ellie, and Louise, have been assigned to a set of rooms consisting of a study room, a small single bedroom, and another small bedroom with a two-person bunk bed. They discuss the proposal that they should take turns, each getting the single for one-third of the school year. Sally refuses to consider this proposal and insists on keeping the single for herself the whole year.

Harman suggests, rightly, that we would regard Sally as unreasonable, though not necessarily irrational.

In the paper from which this example is taken, Harman emphasizes the importance of various distinctions – between theoretical and practical rationality, which he takes to correspond to that between theoretical and practical reasoning (p. 13), between ideal and ordinary rationality (pp. 11-13, 21-3), between epistemic and nonepistemic reasons (pp. 17, 44), between implication and inference (p. 18), between inconsistency and irrationality/unreasonableness (p. 19), and between deduction and induction (pp. 27-32). But except for his comment on the above example, he does not distinguish between 'rationality' and 'reasonableness': indeed, he seems to use the two terms (and their cognates) more or less interchangeably. However, as our initial remarks suggest, cases where we would say that someone was being 'unreasonable' but not 'irrational' are not hard to think up. Refusing to let my children play outside, or not spending enough time with them, or setting someone tight deadlines, or imposing heavy workloads, or perhaps, in general, pursuing 'self-interested' ends, may all be 'unreasonable' rather than 'irrational'. In the first case, I may well – quite 'rationally' – believe that there is a greater chance of my children coming to harm if they are outside alone; yet it is 'unreasonable' in the sense that they are being deprived of experiences – and perhaps, most importantly, the experience of freedom – that are essential to the development of a normal human being. Of course, if I believe that my children should receive as full a range of experiences as possible, then my prohibition counts as 'irrational' and not just as 'unreasonable' – in the same way that Sally would be irrational if she believed that she should do her best to get on with her room-mates.

This does suggest, then, that whilst 'rationality' has more to do with the internal consistency of a set of beliefs, as held by (or revealed by the actions of) an individual at a particular time, 'reasonableness' has more to do with the acceptability of an individual's beliefs in the wider context, such as that of the community. Insofar as the wider beliefs or interests are, after all, shared by the individual, then they would be called 'irrational' and not just 'unreasonable' in the cases we have imagined.

Now some such contrast is just what Harman wants to draw, in emphasizing the distinctions mentioned above. There is one notion of 'rationality' – 'ideal rationality' – which has to do with consistency, and another – 'ordinary rationality' – which has more to do with everyday actions and beliefs. The problem with Harman's account, however, is that it severs the link between rationality and consistency that does seem to be essential to our *ordinary* concept of rationality (as opposed to reasonableness). According to Harman, it may be rational to be (knowingly) inconsistent, but this is not what I think we should say.[1]

The key issue lies in the distinctions Harman wants to draw (pp. 18-19) between (1)

1. I here gloss over the distinction between being inconsistent and being knowingly inconsistent.

and (2), and (3) and (4):

(1) A, B, C imply D.
(2) If you believe A, B, C, you should (or may) infer D.
(3) Propositions A, B, C are inconsistent with each other.
(4) It is irrational (or unreasonable) to believe A, B, C.

On Harman's account, (1) and (2), and (3) and (4), are not equivalent: the first may be true without the second being true (though we should note that (2) cannot be true without (1) being true; the situation with regard to (3) and (4) is more complicated).

The step from (1) to (2), however, is certainly legitimate with the 'may' in brackets replacing the 'should'. For if A, B and C imply D, and we believe A, B and C, then whether or not we are 'rational' or 'reasonable', we surely *may* infer D. But *should* we infer D? Harman's main objection to this is that, since any set of propositions has infinitely many consequences (many of them trivial), it would be most *un*reasonable to expect anyone to infer them all. But if we were to exclude 'trivial' consequences (e.g. by restricting ourselves to 'compact entailments', as defined by Wright 1989), then though it may still be unreasonable to expect anyone to infer all the remaining consequences (think of arithmetic), why would it be irrational? The answer is that it would only be irrational *when set against other beliefs*, i.e. concerning the better things that we may have to do in life. Insofar as it is irrational, then, and not just unreasonable, it is because it is inconsistent with other beliefs.

The issue as to the relationship between rationality and consistency is raised by (3) and (4) directly. Let us agree that if A, B and C are inconsistent, it may not be *unreasonable* to believe all three, for the essential reason Harman gives – we may not know which to reject.[2] Clearly also, if we have rejected the equivalence between (1) and (2) – in its 'should' form – then we have undercut one possible argument for accepting the equivalence between (3) and (4), since to say that A, B and C are inconsistent is to say that they imply a contradiction, but if it is not the case that we *should* infer the contradiction, then it may not be unreasonable to continue to believe all three. But is it irrational? And if we *have* inferred the contradiction, then is it not now, at least, irrational? Again, it may be *reasonable* to continue to believe all three, in the absence of knowing which to reject, but

To claim that it may be rational to be inconsistent seems less controversial than to claim that it may be rational to be knowingly inconsistent. But the contrast is far from straightforward, and both claims are equally problematic. Consider the claim that it is rational to be *consistent*. As a normative judgement, this does seem to presuppose not only that one knows what beliefs one has, but also that one is able to determine whether or not these beliefs are consistent. To suggest instead that it is only rational to be *knowingly* consistent looks distinctly odd – as if one would be better off not knowing whether one is consistent or not, to avoid charges of irrationality. Similarly, to suggest that it may be rational to be *in*consistent (cf. Harman, p. 13), whether knowingly or not, is not what we should say. We can accept that someone can be (knowingly) inconsistent and still act rationally (depending on where the inconsistency is located and what they do), but that is different from claiming that it may be rational to be (knowingly) inconsistent.

2. Note how this presupposes that the inconsistency is recognized. Of course, if it is not, then *a fortiori* it may not be unreasonable to believe all three. This reinforces the point that judgements of reasonableness concern the appropriateness of the relevant behaviour, in this case, the holding of a set of beliefs, in the circumstances.

equally, it would be *irrational* to carry on as if nothing had changed. The judgement of reasonableness reflects an assessment of what is appropriate or understandable in the circumstances, the judgement of irrationality a demand that there be some response to the inconsistency. If it were rational to (knowingly) have inconsistent beliefs, then what motivation would there be to remove the inconsistency? [3]

3. The Paradox of Rational Inconsistency

Distinctions clearly need to be drawn in discussions of rationality; the dispute is over the best way to do so. Harman's account makes it sound as if only the 'ideally rational' person need worry about inconsistency, but this undervalues the normative aspect of our ordinary judgements of rationality. There is certainly something paradoxical about Harman's position. On the one hand we have our ordinary belief that it is irrational to (knowingly) have inconsistent beliefs, yet on the other hand, we have Harman urging us to believe that this is not necessarily irrational. But if he is right, then why can't we believe both? If it is *not* necessarily irrational to be (knowingly) inconsistent, then why can't we also insist that it *is* irrational to be (knowingly) inconsistent? Ironically, it seems, Harman is reluctant to accept higher-level inconsistency. Given the practical benefits and normative role of the belief that inconsistency is irrational, or the difficulty Harman has himself generated in deciding which of these two meta-beliefs to reject, this seems inconsistent. (At the very least, we need an argument to resist the threatened infinite regress.)

Of course, if we distinguish between rationality and reasonableness, and insist on the irrationality but not necessarily unreasonableness of (known) inconsistency, then we face the same paradox with regard to reasonableness. If it can be reasonable to be (knowingly) inconsistent, and we believe this, then why can't we also believe that it is unreasonable to be (knowingly) inconsistent? The answer is that it is sometimes reasonable and sometimes unreasonable, depending on the circumstances (which are what are relevant in our judgements of reasonableness). And if we've admitted that it is *irrational* to be (knowingly) inconsistent, then we've taken care of the normativity that is required here, and we do not also need to hold that it is always unreasonable to be (knowingly) inconsistent.

We started off talking of people being rational but unreasonable, and have ended up talking of people being reasonable but irrational. Judgements of rationality reflect the in-

3. This was the objection that Jan Wolenski raised in discussion to Graham Priest's talk on 'Why It's Irrational to Believe in Consistency' at the Wittgenstein Symposium (2000). Taking Bohr's theory of the atom as the example, Priest replied that it was the empirical inadequacy rather than inconsistency of this theory that prompted its rejection. But whatever the truth of this may be, this response cannot be given in the main case that Priest discussed in the last part of his talk – views about truth. The development of sophisticated theories of truth over the last century has precisely been motivated by the inconsistency in the naïve conception of truth (as revealed by the Liar paradox). Given the complexities of these theories, and the fact that they too often generate contradictions elsewhere, Priest suggested that it was currently rational to retain our (inconsistent) naïve conception of truth. It may be *reasonable* to do so, to the extent that we are unsure which other view to accept, but what is *rational* is to continue to search for a better conception, as indeed many people have been doing.

ternal consistency of a set of beliefs, whilst judgements of reasonableness reflect the acceptability or appropriateness of those beliefs in the specific contexts in which they are held. So there is clearly no straightforward connection between rationality and reasonableness: one does not automatically imply the other. [4] On the contrary, the experience of trade-off here is familiar: the more one sticks to one's guns, to maintain rationality, the more one may become unreasonable; and the more one is reasonable, the more one may have to sacrifice the demands of particular systems of rationality. To say that the ideal person is both rational and reasonable, then, does not properly capture the reflective dynamic here. One should be as rational as possible within the constraints of reasonableness. [5]

4. Rational and Reasonable Reconstructions

One obvious case which calls for the distinction between rationality and reasonableness is the Prisoner's Dilemma. Given the way in which the Dilemma is standardly set up, it does seem that the 'rational' decision is to confess. But clearly, from outside the perspective of rational self-interest, such a decision is unreasonable. (And even if the Prisoner can reckon with the reasonableness of his fellow defendant, it is still unreasonable for him to confess, not least because he would thereby be recognizing the legitimacy of 'reasonable' action.) [6] Now this is not the place to explore how conceptions of rationality can be developed to bring deliverances of rationality in line with those of reasonableness. But it is worthy of note that in his later reflections on the contractarianism he developed in *A Theory*

4. Simple tests of substitutability also demonstrate this. It is rational to be rational, and reasonable to be reasonable, but not necessarily reasonable to be rational or rational to be reasonable. 'It is rational to do *X*' is not equivalent to 'It is reasonable to do *X*'.

5 In his talk at the Wittgenstein Symposium (referred to in n. 3 above), Graham Priest started with the example of the Law of Non-Contradiction. He suggested that Aristotle's failure to provide an adequate proof of this Law meant that it was irrational of him to have accepted it (and given that no-one has subsequently succeeded in proving the Law, equally irrational of us to repudiate inconsistency). I want to say that ungrounded belief in the Law of Non-Contradiction is perfectly *reasonable*, and whether it is *rational* depends on what other beliefs someone consciously holds. This is not the place to engage in exegesis of Aristotle's texts, but it seems clear that Aristotle himself stressed that there could be no such thing as a *proof* of the Law, given its fundamental status in his system. Any proof relies on certain principles, which cannot themselves be proved (within that system). There is only irrationality here if Aristotle also held that everything must be provable; but he didn't. (This is not to say that one cannot engage in a different mode of 'arguing' for the Law of Non-Contradiction – e.g. by showing that anyone who tries to say anything at all *presupposes* the Law. This, I take it, was just what Aristotle attempted to do.) In fact, it seems to me that Aristotle recognized something like the distinction between rationality and reasonableness that I am here concerned to articulate. It is *rational* to believe in whatever can be demonstrated from one's fundamental principles; but it is our sense of what is *reasonable* that tells us which fundamental principles should be accepted. Cf. what is said about Rawls in the next section.

6. Indeed, if he did recognize the legitimacy of 'reasonable' action, then it may be irrational as well as unreasonable for him to confess, since the broader perspective would then be a factor in his own reflections. Once again, this illustrates the subtle complexities of the relation between the rational and the reasonable.

of Justice, John Rawls (1980, pp. 528–30) recognized the need to distinguish between the rational and the reasonable. On Rawls' conception, there are principles of rational choice that guide the negotiation of the agents in the imagined 'original position', but standing behind these principles are principles of reasonableness (e.g. concerning the veil of ignorance) that constitute the framework of cooperation within which the negotiations can proceed. Rawls talks of the reasonable as 'framing' the rational (p. 532); and this seems to confirm what is central to the distinction – that whilst rationality concerns the internal workings of a system, reasonableness has more to do with the wider context in which systems of rationality themselves operate.

That rationality itself must come before the tribunal of reason has recently been powerfully advocated by Wolfgang Welsch (1996: see especially Part II). Welsch emphasizes the complexities of the relationship between reason and rationality (not least with regard to the nightmares of terminology), but argues that the task of 'reason' ('Vernunft') is, firstly, to determine, clarify and adjudicate the boundaries and interconnections between the various forms of rationality, and secondly, to provide a perspective on the whole. What he calls 'transversal reason' ('Transversale Vernunft') is what is needed to judge the transitions and dialectical relationships between forms of rationality. Such a conception may seem a long way from those ordinary intuitions with which we began, but it only further confirms their central message.

With this convergence of ideas from both the 'analytic' and 'hermeneutic' traditions of philosophy in mind (though admittedly with Kant as their common source), perhaps we should conclude by placing the distinction between rationality and reasonableness itself in this wider context. It is no doubt an exaggeration to say that whilst 'analytic' philosophy has tended to focus on 'rationality', 'hermeneutic' philosophy has been more concerned with 'reasonableness'; but exaggerations often serve a useful purpose. Approaches to the history of philosophy can provide the example here. Following Rorty (1984), a distinction is frequently drawn between 'rational reconstructions', which seek to articulate a philosopher's thought as systematically and consistently as possible, from a modern perspective, and 'historical reconstructions', which aim to explain that thought in its actual historical context. But it is more useful here to talk of a contrast between 'rational' and 'reasonable' reconstructions. There is no doubt much value in ironing out the wrinkles in a given philosopher's system (though they always pop out elsewhere, since face-lifts don't last long in philosophy), but equally, justice seems to demand a more 'reasonable' approach, involving sensitivity to the wider context that helps explain the inevitable inconsistencies – inconsistencies which provide the dynamic of philosophical development. 'Rationally reconstructing' a past philosopher's thought seems analogous to a boss who says: "I don't care what you really think, I just want to fit you into my own plan for the department." A plea for reasonableness must accompany all drives for 'rationalization'.

References

Harman, G. 1999 *Reasoning, Meaning, and Mind*, Oxford: Oxford University Press.
Priest, G. 2001 "Why It's Irrational to Believe in Consistency", lecture given at the 23rd International Wittgenstein Symposium (in this volume).
Rawls, J. 1980 "Kantian Constructivism in Moral Theory", *Journal of Philosophy*, 77, 515–72.
Rorty, R. 1984 "The historiography of philosophy: four genres", in R. Rorty, J. B. Schneewind and Q. Skinner (eds.), *Philosophy in History*, Cambridge: Cambridge University Press, 49–75.
Welsch, W. 1996 *Vernunft*, Frankfurt am Main: Suhrkamp.
Wright, C. 1989 "The Verification Principle: Another Puncture – Another Patch", *Mind*, 98, 611–22.

Thomas Hobbes: The Rationalization of Religion

ANAT BILETZKI

> Law: What makes you say, that the Study of the Law is less Rational, than the study of the Mathematicks?
> Philosoph: I say not that, for all study is rational, or nothing worth; but I say that the great Masters of the *Mathematicks* do not so often err as the great Professors of the Law.
> (Thomas Hobbes, *A Dialogue Between a Philosopher and a Student of the Common Laws of England*)

1. Background – Hobbes on Politics and Religion

Famously considered the father of modern political science Thomas Hobbes bequeathed us such staples of political discourse as "the state of nature", "the war of all against all", "homo homini lupus est", the "social contract", an absolute monarch, and the concept of "rights".[1] Less famously, but more perplexing for scholars, is the fact that Hobbes devoted half of his masterpiece, the *Leviathan*, [2] to questions of religion – religious belief, religious authority, and religious interpretation.

So it is – or was, for a long time – almost trite to claim that Thomas Hobbes's *Leviathan* is really two books: one political, the other theological. Were his political theory less aggressively "scientific" this duality would not pose such a blatant problem. But it does become enigmatic given the following theoretical tension: how can one countenance the logical, methodical, analytic designs of the book that purports to supply humankind with a *science* of politics (including definitions of rights, law, and sovereignty), concomitantly with the religious, theological presentation of a very "Christian Commonwealth"? Around the issue of Hobbesian theology the scholar becomes an apologist. Indeed, Hobbesian exegesis is intensely employed in bridging this apparent inconsistency, or at least explaining it.[3]

The plurality of explanations for this puzzle notwithstanding, there are, in the main, two strategies of viable (and traditional) solutions;[4] call them the religious strategy and

1. This is not to say that he was the first to use these terms, nor to say that these were his exact words; only to say that they are forever associated with his name.
2. And two chapters in the *Elements of Law*, and the third part, "Religio", of *De Cive*.
3. For a comprehensive presentation of the interpretive issues arising from Hobbes's religious writings see King (1993).
4. And there is, or was, of course, the view held by (most of) Hobbes's contemporaries: that there is no "solution"; that is to say, that Hobbes must be understood, and castigated, as pretending to be theistic in the second half of *Leviathan*, but failing to cover up his political atheism. Erstwhile representatives are: Ross 1653, Lawson 1657, Lucy 1663, Tenison 1670, Clarendon 1676, Bramhall 1677, Whitehall 1679, Parker 1681.

B. Brogaard, B. Smith (Hg.), *Rationality and Irrationality / Rationalität und Irrationalität*, S. 25–35.

the political strategy. The former takes Hobbes at his word – his religious words, that is. Such exegetes "believe" Hobbes when he avows his Christian faith, and regard books III and IV of *Leviathan* as true proclamations of his ideology of grounding a *Christian* commonwealth. Verily, since Hobbes's status as philosopher of politics is well-established, this type of Hobbesian reading must address the problem of sovereign power vs. religious authority and, indeed, we find in the literature a plethora of answers. Some, from very early on, honestly recognize the inevitable problematics, and leave things as they are: i.e., problematical. A poignant example is John Hunt, who admits, in 1870: "It is difficult ... to reconcile Hobbes with himself ... it is certain that he did ascribe to his grotesque monster a power to make right and wrong, and to dictate both religion and laws to the people. This position, even as laid down by Hobbes himself, seemed to leave no other foundation for either religion or morality than the will of the sovereign" (Hunt, 410). Others sometimes see the religious parts grounding the political, sometimes the ethical grounding the religious, in all cases accepting the final identity between "natural law" and God's commands.[5] And among these others there are exegetes who investigate the ins and outs of the second half of *Leviathan*, the religious writings, to scrupulous in detail, addressing, e.g., the differences between "standard" Christian and "orthodox" Christian, or between Catholicism, Protestantism and Calvinism, all with the aim in mind of grounding Hobbes's theism.[6]

The latter, political strategy, is, to my mind, more an external, contextual explanation of the above tension than an analysis of the in-Hobbes problematics involved, at least to begin with. The apparent inconsistency in Hobbes is sometimes almost shrugged away with an excuse for Hobbes, stating that Hobbes had no choice but to admit a religiosity conforming with the powers that be, given his times. Here Hobbes is seen as anything from apologetic to opportunistic. Early on, but continuing throughout the saga of Hobbesian interpretations, interpreters recognize the "twofold truth, philosophical and theological" (Lange 1881, 218) visited on early philosophers of modernity, and acquiesced to in Hobbes's own words: "The subject of philosophy ... excludes *Theology* ... the doctrine of *angels*, and all such things as are thought to be neither bodies nor properties of bodies" (*De Corpore*, 1.8). But, looking at Hobbes dealing with religion in the second part of *Leviathan*, and viewing it as "rhetorical, ironic, and prudent," (Cooke 1996, 18) or as "artful equivocation," (Oakeshott 1962, 283) most interpretations of this bent stay aware of the internal, radical incompatibility between the political and the theological.[7] Some contribute, like in the case of the religious bunch, detailed and stalwart exegesis purporting to explain Hobbes's atheism, in spite of the second half of *Leviathan*, and to give a cohesive account of the historical-ideological context of his rhetorical biblical interpretation.[8] Also, on the meta-level, as it were, there is the interpretation that sees Hobbes as intentionally ambiguous, intentionally thwarting of clear understanding; a manipulative Hobbes who knew what it was that he "intended to remain covered; that the formerly highest authorities – the Bible and God – have been dethroned and replaced by the sovereignty of man" (Cooke 1996, 36-7). But this is still an "external" strategy, although, admittedly, one buttressed by serious, exegetical, "internal" argument. And, indeed, interpreters in this "political" camp talk of "secularization" (Mintz 1962)[9] and "rationalization" (Campdonico 1982).[10]

5. Maurice 1862, Dewey 1918, Taylor 1908, Warrender 1957.
6. See, e.g., Hood 1964, Goldsmith 1966, Pocock 1973, Johnson 1974, State 1991, Wright 1991, Martinich 1992, and others.
7. See Willey, 1934 and Strauss, 1950.
8. See, especially, Curley 1992.

Still, neither strategy, religious or political, grasps the bull by his horns, for the inconsistency is a profound and real contradiction. One cannot insist on a mechanical, materialistic ontology coupled with a rationalistic, contractarian view of social and political authority, while sincerely holding on to a supreme, all-powerful, omniscient God behind Holy Scripture. And the contradiction runs rife with the various political and scientific interpretations extant of Hobbes. Mechanical materialism of the scientifically deterministic type cannot sit well with religious, godly determinism.[11] A "conservative" view of Hobbes, as espousing an absolute ruler who establishes morality, is similarly at odds with God-given moral commands; and the (now in-fashion) liberal reading of Hobbes, as upholder of a rights-based view of man, is just as inconsistent with a religious position that posits God as the authority behind ethics.

2. More Background – Hobbes on Language

It is relatively new – the last 30 years or so – to point to the central role that language plays in Hobbes's general philosophical thought – to the fundamental status of speech in his analysis of man. Needless to say, there is no one accepted interpretation concerning his thought on language or, even less, concerning the relations which exist between his theory of language and thought on the one hand and political or religious theories on the other. There is a standard interpretation of his philosophy of language, which finds its anchor in several texts and contexts, as being denotational and representational. But a discussion of generally neglected aspects of Hobbes's philosophy of language shows it to be a theory of language use that has outstanding implications for his moral philosophy. These Hobbesian views focus on man making use of language, beyond language being a semantic, denotational, representational, referential system or what not. Encountered in *Leviathan*, in the later, more "scientific" *De Corpore*, and most explicitly in the earlier *Human Nature*, language is defined via its use. A further insight, that we do things with language, lays the foundation for an active and creative view of language. This theory of language which is ruled by use may be seen as expressing the conditions for speech, as supplying the starting point from which to look at meaning, and as expanding the range of language to encompass different spheres of human behavior. So it is precisely this view which conduces to a language-oriented gaze at morality. Speaking of morality, Hobbes recognizes two systems of laws within the sphere of morals: the laws of nature and the laws of society. His originality consists of finding a new and different rationale for the so-

9. This label of "secularization" is not to be confused with Martinich's (1992) distinction between secular and religious interpretations, where the first are those that do not credit religious concepts with true significance in Hobbes's thought.
10. Even as early as 1874 (Tulloch) there abounds the realization that one can talk of "rational" theology. And, in a sense, one may point to all theology (since Aquinas) as a rationalization of religion. This is not our point in ascribing a rationalization of religion to Hobbes.
11. But see the integration of Hobbes's materialism and his theology in Pacchi 1988.

cial law other than (but not contradictory to) the laws of nature. This is done by positing the important connection between language and society. Both are human artifacts; both are constituted by man; and one cannot be countenanced without the other. The sovereign, by the very act of legislating, gives moral terms meaning within the context of a society, and thus institutes social law. This meaning-giving act is constitutive, and labeling it so is crucial in explaining the move from the state of nature to the social state. Social and linguistic constructs attain meaningfulness ("during" the move from the state of nature to the social or linguistic state) only through constitutive rules. The sovereign, thus, in enacting laws, defines, or constitutes, the (societal) meaning of "just", and all other moral and political words.

3. Language to the Rescue

Now, after these background noises, I should like to attempt a novel type of dissolution of our contradiction – between two manifestations of Hobbes – by turning to language and rationality.[12] This is done by investigating the problem of the authority of Scripture which is explicit in Hobbes's description of such authority. Hobbes himself is aware of the incongruities between his demands for rational criteria of knowledge and the status of scriptural knowledge. Recognizing such incongruities has been in the realm of the political strategy outlined above. That is, it seems that the only way out of such internal contradiction is external excuses for the contradiction. We shall, however, travel a different route by positing Hobbes's sovereign as a meaning-giving authority;[13] and we shall show that such understanding of both Hobbes's philosophy of language and his theory of political authority can make sense of the seeming religious incongruities. For, given the sovereign's authority to confer meaning upon words, it is no wonder that he is invested with the right and duty to interpret Holy Writ. Interpretation then becomes the key to our puzzle.

The route of Hobbes's thought in the *Leviathan* must be surveyed – beginning with his definitory politics (of rights, duties, and laws) and leading to religion and religious interpretation. There is a strain between, and not altogether consistent form and content of, different layers of the writings: explicit credos asserting liberty and equality, methodological attempts at a scientific building of political rights, imaginary stories of states of nature and universal meetings, systematic and explicit utilitarian interpretations of Scripture, etc. Hobbes's writings on rhetoric, on law, on logic and on language (in texts other than the *Leviathan*) can also be employed for their social and scientific underpinnings of his thought on religion. And in the specific writings on religion one can hope for a breakthrough by moving from religious *belief*, to the all-important function of religious *interpretation* and its related derivation of religious *authority*. The place of Scripture, the existence or status of God, and the obligations of a pious life can all be made sense of through the vehicle of interpretation. For interpretation, in Hobbes, is rule-governed and

12. This perspective has been addressed by Rhodes (1989), where an explanation of the religious parts of *Leviathan* is proferred as an application of the theory of the politcial parts through acknowledgement of the linguistic aspects of the latter.
13. See Biletzki (1997) for an analysis of Hobbes's pragmatic theory of language – in social, contextual, and functional terms.

authoritative, i.e., scientific and political. Furthermore, when the authority to interpret religious writings is given to the sovereign – i.e., the political sovereign – religion ceases to be divorced from politics at the very pinnacle of theoretical thought about them both.

If, then, his politics is a "rational" politics one must make sense of this precise rationality as fitting in with religion and Hobbes provides a solution to the problem of "fitting in" through the interpretative authority of the sovereign. One can go further here: religion constructs irrational parochialism by abhorring logic, language and analysis. If logic, language and analysis be deemed rational then granting the sovereign the right to interpret Scripture (i.e., to use logic, language, and analysis) and thereby to embody religious authority is precisely the move we're looking for: the onus of rationality at the hands of (political) sovereignty.

4. Interpreting Scripture

In pondering the place of Scripture – i.e., its authority, its status, and finally its interpretation – we may turn to numerous sections in Part III of *Leviathan: Of A Christian Commonwealth*, and to its end, where Hobbes claims he has given us all that is needed for "Policy Ecclesiasticall". The perils of exegesis are such that one may find an evidential quote for almost any position, and, moreover, that it is precisely Hobbes's massive verbosity that has given rise to the tensions mentioned above. Blatantly, perhaps even superficially, one may say that this part of Leviathan explicitly extols the authority of Scripture, on the one hand, while limiting its power, on the other. Suffice to emphasize that this two-edged attitude towards Scripture within a context which deals with Scripture is the source of our profound problematics. Had Hobbes merely added on to the first two parts of Leviathan (*Of Man*, *Of Commonwealth*) an additional "religious" part, the task of understanding him would have remained, indeed, an external one and it would have been plausible to accept the political strategy outlined above with ease. A deeper problem is encountered when, within such religious writing, one meets two seemingly opposing views on Scripture itself. Its solution (and my case) will rest on showing that Scripture is, indeed, at the hands of the sovereign.

Famously, Hobbes incurred the wrath of Church, "believers", and all religious institutions by placing his sovereign at the apex of power, and thereby placing Scripture below that sovereign.

> By the Books of Holy SCRIPTURE, are understood those, which ought to be the *Canon*, that is to say, the Rules of Christian life ... Seeing therefore I have already proved, that Soveraigns in their own Dominions are the sole Legislators; those Books only are Canonicall, that is, Law, in every nation, which are established for such by the Soveraign Authority. (*Leviathan*, 33)

Naturally, then, deciding *what* is Scripture is up to the sovereign. Why not, however, place God above all sovereigns, as sovereign of sovereigns, and by so doing solve both the internal problem of the place of Scripture and the external wrath of official religious powers. Hobbes ruminates on this possibility and encounters a dead end.

> It is true, that God is the Soveraign of all Soveraigns; and therefore, when he speaks to any Subject, he ought to be obeyed, ... But the question is not of obedience to God, but of *when*, and *what* God hath said; which to Subjects that have no supernaturall revelation, cannot be known, ... (*Leviathan*, 33)

Subsequently, the problem of the place of Scripture cannot be solved without a turn to the problem of *interpretation*; for even if we speculate a God above all sovereigns we have no accessibility to his word save by interpretation. Accordingly, we must ask of the authority to interpret and of the rules of interpretation. Let us address the latter first.

5. Rules of Interpretation

Two prima facie rules of interpretation seem to place man – ordinary man, not necessarily a sovereign – and his natural talents and propensities at the steering wheel of interpretation. Reading the word of God must be constrained by two human faculties which had played the pivotal role in the theory of knowledge expounded in *Of Man*: the senses and reason.

> Nevertheless, we are not to renounce our Senses, and Experience; nor ... our naturall Reason ... For though there be many things in God's word above reason; that is to say, which cannot by natural reason be either demonstrated or confuted; yet there is nothing contrary to it; ... (*Levaithan*, 32)

Yet, lest we be seduced into thinking that Hobbes is unambiguously placing human criteria above the godly, Hobbes immediately, within the same passage, explains the status of such human faculties.

> For they are the talents which [God] hath put into our hands to negotiate. (*Leviathan*, 32)

Is this mere lip service?

Interpretation is a hermeneutic art, and Hobbes is well-aware of that. Were reading the Scriptures limited to having them agree with our sensual experience and logical reasoning, the reader of Scripture would obviously face insurmountable difficulties which could not be explained by "our unskillful Interpretation, or erroneous Ratiocination" alone. So Hobbes lays down, or rather uses, a further rule for interpretation: a distinction between plain (or proper) and metaphorical readings. He has a preference for literal readings, founded on his well known dislike of metaphors. The famous diatribe against metaphors is also a well known metaphorical tract:

> The Light of humane minds is Perspicuous Words, but by exact definitions first snuffed, and punged from ambiguity; *Reason* is the *pace*; Encrease of *Science*, the *way*; ... And on the contrary, Metaphors, and senslesse and ambiguous words, are like *ignes fatui*; and reasoning upon them, is wandering amongst innumerable absurdities; (*Leviathan*, 5)

Indeed, in several places, he manages to derive and explain a literal interpretation calling it "proper" even in opposition to the received one. Nevertheless, even the literal minded Hobbes cannot escape an occasional metaphorical interpretation (as opposed to his own metaphorical *style*), in places where "proper" reading would become ludicrous. Thus, for instance:

> On the signification of the word *Spirit*, dependeth that of the word INSPIRATION; which must either be taken properly; and then it is nothing but the blowing into a man some thin and subtle aire, or wind, in such manner as a man filleth a bladder with his breath ... That word therefore is used in the Scripture metaphorically onely ... (*Leviathan*, 34)

Yet, most significant for our purposes is a further rule – the retroactive rule of interpretation which Hobbes preaches in the summary of Part III. Ignoring the plain/metaphorical distinction above, Hobbes turns rather to the plain/obscure distinction as that which can be interpreted versus that which cannot. Here, however, obscurity, leading to conflicts of interpretation, does not arise from the words themselves (which may still be proper or metaphorical) but from a misunderstanding of the art of reading. Proper reading must take context into consideration; context in the sense of the writer's circumstances, context in the sense of the overall text, and context in the sense of the writer's aims and intentions.

> And in the allegation of Scripture, I have endeavoured to avoid such texts as are of obscure, or controverted Interpretation; and to alledge none, but in such sense as is most plain, and agreeable to the harmony and scope of the whole Bible; which was written for the re-establishment of the Kingdome of God in Christ. For it is not the bare Words, but the Scope of the writer that giveth the true light, by which any writing is to bee interpreted; and they that insist upon single Texts, without considering the main Designes, can derive no thing from them cleerly; but rather by casting atomes of Scripture, as dust before mens eyes, make every thing more obscure that it is; an ordinary artifice of those that seek not the truth, but their own advantage. (*Leviathan*, 43)

6. The Authority to Interpret

Given that interpretation must accord with our senses and our reason, that it must properly construe words as plain or metaphorical, and that it must account for context, the question of the authority to interpret still looms large. For, even if we obey such rules of interpretation, we may encounter different interpretations. Furthermore, on a meta-level we may even ask about the rules themselves: whose is the authority to constitute rules of interpretation? And Hobbes gives us an inkling of his answer while leaving the question, for the meantime, open:

> Which question cannot bee resolved, without a more particular consideration of the Kingdome of God; from whence also, wee are to judge of the Authority of Interpreting the Scripture. For, whosoever hath a lawfull power over any Writing, to make

it Law, hath the power also to approve, or disapprove the interpretation of the same. (*Leviathan*, 33)

Ruminating over the pitfalls of ecclesiastical writings and their interpretation, and hesitating over the point he would like to make, Hobbes invigoratingly goes into minute details of both the Old and the New Testament, and finally points to his solution:

> ... seeing the Examination of Doctrines belongeth to the Supreme Pastor, the Person which all they that have no speciall revelation are to beleeve, is (in every Common-wealth) the Supreme Pastor, that is to say, the Civill Soveraigne. (*Leviathan*, 43)

Such formulation (i.e., Supreme Pastor) of interpretative authority runs the risk of being understood as a compromise between religious and civil authority. Yet there are clear hints of Hobbes's insistence on the priority of civil powers in his actual divorce of God and specific Christian God, telling us that "Obedience to God and to the Civill Soveraign" is not "inconsistent, whether Christian, or Infidel" (*Leviathan*, 43).

Rather than further rehash the standard Hobbesian questions of authority and interpretation, and even re-affirm the understanding of Hobbes which places him at the civil rather than religious end of the spectrum, we can substantiate this way of looking at Hobbes by turning to his thought on ("philosophy of") language and, not independently, to his political and moral philosophy. To elucidate: one can better understand what Hobbes had to say about religion, religious law, God, and religious writing – and one can more strongly claim that what he had to say was not standardly "religious" – by investigating his positions in the light of his philosophy of language.

Another way of looking at this would be to say that we're turning Hobbesian exegesis upside down. Even those interpreters of Hobbes who agree with the "anti-religious" reading of Hobbes see Part III of *Leviathan* as a (somewhat cynical) attempt to ground his political and social theories in choice scriptural writings. But Part III is fully in accordance with, and indeed an instance of, Hobbes's unique view of language which makes for a different understanding of his political and moral philosophy. Only by incorporating Part III into the whole of *Leviathan*'s theory of meaning – the meaning of all terms, whether natural, moral, social, or religious – can one get a consistent reading of Hobbes's philosophy of religion.

7. Conclusion

Hobbes amazingly, in the seventeenth century, offered us, beyond a political theory generally recognized as revolutionary (positively or negatively), a linguistic theory far ahead of the semantic standards of his times, and (in consonance with these two) a look at Scripture more convoluted than others of his times.

Put simply, one can formulate the tensions of Part III of *Leviathan* as God versus Sovereign, and Sovereign's authority versus scriptural authority. Any way of labeling the tension leads to the problem of interpretation and Hobbes, in fact, has focused the problematics by claiming that the author of interpretation is the sovereign.

> It may therefore be concluded that the interpretation of all laws, as well sacred as *secular* (God ruling by the way of *nature* only), depends on the authority of the city, that is to say, that man or counsel to whom the sovereign power is committed. (*De Cive*, XV, 17)

Yet such a focus does not solve the problem; it merely highlights Hobbes's (conscious or otherwise hidden) preferences and sharpens our question marks.

If, however, we ask not of the sovereign's political authority, but of his linguistic authority (or, indeed, claim that the one inheres in the other), we get a less arbitrary characteristic of his power. Authority, indeed, becomes a part of linguistic practice, and, conversely, linguistic acts become profoundly social, even political. However, in order to clearly ask such questions one must adopt, and attribute to Hobbes, a becoming theory of language: one that explicitly and implicitly recognizes language as language in action.

Treating of language Hobbes turns to what we would anachronistically call "speech acts". Speaking of morals, Hobbes has the sovereign (and we, anachronistically, have Hobbes) performing performatives. Perusing interpretation, Hobbes leans fundamentally on context.

If such anachronisms be valid we may conclude: Hobbes saw language as an activity through which meanings are given by use and accredited the sovereign with power over giving meanings in all social contexts. Religion being such a context, it is in the sovereign's hands to confer meaning upon religious terms. Interpretation then becomes an offshoot of these activities.

But interpretation must be active interpretation. Nowhere is this more clearly or beautifully espoused than in Hobbes's cry to interpret the word *word* itself:

> When there is mention of the *Word of God*, or of *Man*, it doth not signifie a part of Speech, such as Grammarians call a Nown, or a Verb, or any simple voice, without a contexture with other words to make it significiative; but a perfect Speech or Discourse, whereby the speaker *affirmeth*, *denieth*, *commandeth*, *promiseth*, *threateneth*, *wisheth*, or *interrogateth*. (*Leviathan*, 36)

A theory of language in use tells us that we do many things with language. It is the vehicle of communication, it is the media of representation, it is the tool of promises, it is the home of prayer. But we must give fresh answers to the problem of interpretation at large, and religious interpretation in particular, by, paradoxically perhaps, stressing the part of the human – rather than godly – user of Scripture. Hobbes teaches rules of scriptural interpretation utilizing extra-religious constructs for both his own interpretation of Scripture and the sovereign's authority to interpret. Interpretation truly and ultimately becomes a rational "Policy Ecclesiasticall".

Literature

Biletzki, A. 1997 *Talking Wolves Thomas Hobbes on the Language of Politics and the Politics of Language*, Dordrecht: Kluwer.

Bramhall, J. 1677 "The Catching of Leviathan", in *Collected Works*, vol. III, Dublin.

Campodonico, A. 1982 "Secularization in Thomas Hobbes's Anthropology", in J. G. van der Bend (ed.), *Thomas Hobbes, His View of Man*, Amsterdam: Rodopi.

Clarendon, E. 1676 *Brief View and Survey of thee Dangerous and Pernicious Errors to Church and State, In Mr. Hobbes's Book, Entitled Leviathan*, Oxford.

Cooke, P. D. 1996 *Hobbes and Christianity: Reassessing the Bible in Leviathan*, Lanham: Rowman & Littlefield.

Curley, E. 1992 "'I durst not write so boldly', or How to read Hobbes' theological-political treatise", in D. Bostrenghi (ed.), *Hobbes e Spinoza, Scienza e Politica*, Naples: Bibliopolis.

Dewey, J. 1918 "The Motivation of Hobbes's Political Philosophy".

Goldsmith, M. M. 1966 *Hobbes's Science of Politics*, New York: Columbia University Press.

Hood, F. C. 1964 *The Divine Politics of Thomas Hobbes*, Oxford: Clarendon Press.

Hunt, J. 1870 *Religious Thought in England From the Reformation to the End of the Last Century*, Vol. I., London: Strahan & Co. Rpt. AMS Press, 1973.

Johnson, P. 1974 "Hobbes's Anglican Doctrine of Salvation", in R. Ross, H. W. Schneider, and T. Waldman (eds), *Thomas Hobbes in His Time*, Minneapolis: University of Minnesota Press.

King, P. (ed.) 1993 *Thomas Hobbes: Critical Assessments*, vol. IV: *Religion*, London and N.Y.: Routledge.

Lange, F. 1881 *History of Materialism*, Vol. I., Boston: Houghton, Mifflin.

Lawson, G. 1657 *An Examination of the Political Part of Mr. Hobbs His Leviathan*, London.

Lucy, W. 1663 *Observations, Censures and Confutations of Notorious Errors in Mr. Hobbes His Leviathan*, London.

Martinich, A. P. 1992 *The Two Gods of Leviathan: Thomas Hobbes on Religion and Politics*, New York: Cambridge University Press.

Maurice, F. D. 1862 "Thomas Hobbes" in *Modern Philosophy*, London: Griffen, Bohn.

Mintz, S. 1962 *The Hunting of Leviathan*, Cambridge: Cambridge University Press.

Oakeshott, M. 1962 *Rationalism in Politics*, London: Methuen.

Pacchi, A. 1988 "Hobbes and the Problem of God", in G. A. J. Rogers and A. Ryan (eds.), *Perspectives on Thomas Hobbes*, Oxford: Clarendon Press.

Parker, S. 1681 *A Demonstration of the Divine Authority of the Laws of Nature and the Christian Religion*, London: R. Royston and R. Chriswell.

Pocock, J. G. A. 1973 "Time, History, and Eschatology in the Thought of Thomas Hobbes", in *Politics, Language and Time*, New York: Atheneum.

Polin, R. 1981 *Hobbes, Dieu et les Hommes*, Paris: PUF.

Rhodes, R. 1989 "The Test of Leviathan; Parts 3 and 4 and the new interpretations", in M. Bertman and M. Malherbe (eds.), *Thomas Hobbes; De la Metaphysique a la Politique*, Paris: J. Vrin.

Ross, A. 1653 *Leviathan Drawn out with a Hook, or Animadversions Upon Mr. Hobbs His Leviathan*, London.

State, S. A. 1991 *Thomas Hobbes and the Debate Over Natural Law and Religion*, Hamden: Garland.

Strauss, L. 1950 *Natural Right and History*, Chicago: Chicago University Press.

Taylor, A. E. 1908 *Thomas Hobbes*, Port Washington, N.Y.: Kennikat Press.

Tenison, T. 1670 *The Creed of Mr. Hobbes Examined; in a Feigned Conference between Him, and A Student in Divinity*, London.

Tulloch, J. 1874 *Rational Theology and Christian Philosophy in England in the Seventeenth Century*, Edinburgh: Blackwood.

Warrender, H. 1957 *The Political Philosophy of Hobbes: His Theory of Obligation*, Oxford: Clarendon Press.

Whitehall, J. 1679 *The Leviathan Found out: or the Answer to Mr. Hobbes's Leviathan, in that which my Lord of Clarendon hath Past Over*, London.

Willey, B. 1934 *Seventeenth Century Background*, London: Chatto and Windus.

Wright, G. 1991 "Introduction, 1668 Appendix to *Leviathan*", *Interpretation* 18.

Elusive Reference

BERIT BROGAARD

1. Many, But Almost One

Our ordinary uses of ordinary singular terms in ordinary contexts are marked by vagueness. Where, for example, should we draw the boundary around what we call 'the sun'? There does not seem to be only one such boundary, not least for the reason that the sun is continuously changing and continuously losing and gaining particles. We can say, therefore, that there is no single answer to the question of where the boundary of the sun should be drawn.

This is not because we do not know where the boundary lies. Nor is it because there are entities, bits of physical reality, that neither belong nor do not belong to other bits of physical reality. Rather, many aggregates of matter deserve the name 'the sun'. Each has a sharp boundary, but no one aggregate is more deserving of this name than any other. A singular term like 'the sun' thus stands to those aggregates in a reference relation that is one-to-many rather than one-to-one.

Unger (1980), Lewis (1993), and others have already observed that there are many equally good boundaries for 'cloudy' objects such as the sun, this cloud, or that pile of sand. But, as they point out, the problem of the many equally good boundaries is not merely a problem for such blurry entities; it is a problem for almost any one of the common-sense entities that surround us in our everyday lives:

> There are always outlying particles, questionable parts of things, not definitely included and not definitely not included. So there are always many aggregates, differing by a little bit here and a little bit there, with equal claim to be the thing. We have many things or we have none, but anyway not the thing we thought we had. (Lewis 1993, pp. 164-5)

Unger has argued that entities – such as your cat Bruno, the sun, or this cloud – do not exist. For, as he says, if there is not one single candidate for the reference of the corresponding terms, then we might as well say that there is no candidate: having many candidate referents of our ordinary terms is for him a genuine absurdity. Lewis, on the other hand, argues that such entities do exist, but that when we use a term like 'the sun', then we do not ordinarily succeed in picking out any one of the many aggregates deserving of this name.

In the case of 'the sun', we do not need to feel too uncomfortable about the fact that there is no one aggregate that qualifies as its extension. For as Lewis (1993, p. 178) points out, the many aggregates of the sun are not entirely distinct. They are not disjoint mereologically; rather, they overlap. Although no two of them are identical, any two of them are *almost identical*: they share almost all their parts in common. They are many, but almost one.

B. Brogaard, B. Smith (Hg.), *Rationality and Irrationality / Rationalität und Irrationalität*, S. 36–45.

So we can feel a bit more comfortable. We are not really choked by the news that there is no one aggregate that qualifies as the extension of 'the sun'. For it is almost as if there were just one such aggregate. Unfortunately however this comforting fact cannot be generalized. Suppose you are invited to a party at Fred's house. As Lewis points out, 'it never crossed your mind to think whether by "house" you meant something that did or that didn't include the attached garage' (1993, p. 172). 'Fred's house' (even leaving aside the fact that the corresponding buildings themselves are many but almost one in the sense described above) refers to two things that are far from being almost identical: they refer to the house with and to the house without the garage. Or to make matters worse: 'buy one, get one free' is in Buffalo supermarkets often printed above the shelf containing Marlboro cigarettes. The singular term 'the pack I bought' then stands in the reference relation to entities which, far from being almost identical, are in fact entirely distinct.[1]

Lewis sees Fred's house as a genuine example of vagueness in the sense that there are utterances about Fred's house whose truth-value cannot be determined. I shall here defend the view that, even in spite of the vagueness of our singular terms, most utterances involving such terms can nonetheless be assigned a truth-value in fully determinate fashion.[2]

2. Supervaluation

We may attempt to solve the problem of the many sun aggregates by making a decision. The best physicists of the world could gather together and decide where to put the boundary of the sun. In making their decision they could consider such things as the fact that no two heavenly bodies overlap. If we have two things, such as the earth and the sun, then we just know intuitively that this is so – they do not have any parts in common. So at the very least the sun should not be so big as to include particles of the earth. And it should have a certain minimum size as well. It cannot be reduced to the size of a point and still be what we call 'the sun'. After having considered such relevant intuitions, the physicists could make a final decision. They would thus have *precisified* the concept under consideration. But then again, their decision would still be to a high degree arbitrary. They might as well have made another decision which would have complied equally well with our intuitions about objects such as the sun. It turns out that there are infinitely many decisions that could have been made.

The solution to the problem of the many begins by noticing that there are many sentences which would come out true no matter which decision or precisification is in fact made. 'The sun is a star' is true no matter which one of the aggregates is taken to be the referent of 'the sun'. Similarly for 'The sun is beautiful in the evening', 'The sun is essential for all life on earth', 'The sun can cause skin cancer' and so on. This observation constitutes the core of van Fraassen's method of supervaluation. On this method, an utterance is supertrue if and only if it is true (and superfalse if and only if it is false) under all

1. This example is taken from Sorensen 2000.
2. Parts of this paper are based on the theory of truth and reference worked out in greater detail in Smith and Brogaard 2001.

precisifications. If, on the other hand, it is true under some ways of precisifying and false under others, then it falls down a supertruth-value gap. Thus it is supertrue to say: 'the sun is the only star in our solar system' or 'the sun is beautiful in the evening'. These utterances are true no matter which of the many sun aggregates you put into the extension of the term 'the sun'. It is superfalse to say that 'the sun is identical to the earth' or 'the sun is in an orbit around the earth' because these utterances are false no matter which one of the many sun aggregates you put into the extension of the term 'the sun'. Finally, 'this outlying particle is part of the sun' is neither supertrue nor superfalse, since it is true on some precisifications and false on others. Its truth-value is indeterminate.

3. Supervaluation Contextualized

Imagine, now, that you are looking at Mont Blanc from a distance. It seemingly has a sharp boundary that separates it from the surrounding sky; for you cannot see the people and trees on the mountain or the small rabbits crawling around under its bushes. You know perfectly well that there are such things. But in these circumstances you ignore them. Your perception does not separate out the things you are seeing from the things you are ignoring.

Suppose someone asks you, now, whether you think that rabbits are part of Mont Blanc. This very question establishes a new context.[3] The lazy diffuseness of your earlier perceptual projection is hereby brought to an end. For in responding to this question, your use of the term 'Mont Blanc' picks out only aggregates that don't include rabbits as parts.

I would like to propose a way of dealing with these matters which might be called 'reference contextualism' – a view similar to the epistemological contextualism defended by Lewis in "Elusive Knowledge" (1996).[4] The term 'Mont Blanc' singles out mountain aggregates that include rabbits in one context but not in another. Of course, if you begin to reflect on the concepts *mountain* or *Mont Blanc* independently of your present visual experiences, then you would not normally use the concept *Mont Blanc* in such a way that the corresponding object would include rabbits as parts. This is because you know that Mont Blanc is a mountain, and mountains are not normally such as to include rabbits as parts.[5] Thus, when you are thinking critically about how to use the term 'Mont Blanc', you are in effect moving into a context in which rabbits and mountains are treated as disjoint entities. The point is, however, that you would not always use the term 'Mont Blanc' after having reflected upon the term. Or in other words, how you understand the term 'Mont Blanc' will not always affect in the same fashion the way you use it (which means: what families of precisified parcels of reality are its candidate referents). You might as well have used the term in a way that is based on your diffuse and lazy perception. Note also that there could have been contexts in which you would use the term 'Mont Blanc' consciously to include rabbits as parts. Think of a hunting context, or a context in which

3. Searle has suggested that we take context to be somewhat similar to what he calls 'background assumptions' in his *Intentionality*.
4. This view is defended at greater length in Smith and Brogaard (2001).
5. See Fine 1975.

there is a real estate market for mountains including their animals.

It seems, then, that the sentence

[A] Rabbits are part of Mont Blanc

is a clear example of a sentence which falls down a supertruth-value gap, for it seems to be true under some precisifications, but false under others. It seems to be true under the precisification made in the hunting context ("these rabbits are part of my mountain"). It also seems to be true in the context determined by the precisification you make when you are looking at Mont Blanc from a distance. It seems to be false, on the other hand, in the normal context in which you are close enough to Mont Blanc to see the rabbits under its bushes.

The key idea of reference contextualism, however, allows us to see that [A] does not fall down a supertruth-value gap after all. The illusion that it does follows from the fact that it has been assumed that sentences are the relevant entities to look at for purposes of evaluation and that such evaluation should accordingly be effected independently of the context in which sentences are used. This assumption is indeed standard among proponents of supervaluationism. Morreau, for example, suggests that vagueness – for example the vagueness involved in a sentence such as 'Al is bald' arises where there occurs a 'coupling [of] a vague adjective with the name of a borderline case' (1999, p. 149). Such random couplings may be common in Morreau's world. In the world I live in, however, we take care to see that our sentences are seriously and sincerely judgeable in whatever happen to be the relevant contexts. Consider a context in which you are looking at Mont Blanc from a distance. In such a context you could not formulate an utterance such as [A], since this would require that you have first moved to a new context in which rabbits are cognized as separate entities. In this latter context, however, the sentence in question is simply superfalse. This is why [A] seems, in our linguistic community, to be obvious nonsense. It expresses a mereological counterpart of falsehoods concerning identity, such as 'Julius Caesar is a cardinal number' or 'The Morning Star is different from the Evening Star' or 'Chisholm is Searle'. All such sentences are (in the contexts we currently share) not sincerely judgeable. We can, certainly, mouth the corresponding words; but the sentences in question are unjudgeable nonetheless.

Judgeability is a socio-psychological notion. Whether or not something is judgeable depends in part on the linguistic community to which the speaker belongs; a sentence is judgeable in a given context just in case the speaker would feel comfortable sincerely judging that sentence in that context. Normally people would not feel comfortable judging the sentence 'Rabbits are parts of Mont Blanc'. But in the linguistic community of hunters it might be that you would feel perfectly comfortable judging this sentence. The sentence in question falls down a supertruth value gap in none of these (naturally occurring) contexts. Rather, in each of them, it is determinately either supertrue or superfalse.

4. Elusive Reference

Reference hereby becomes elusive in a way that is similar to the elusiveness of knowledge as conceived by Lewis. Knowledge that *p* is elusive, for Lewis, if the very fact that one begins to discuss what possibilities there are that *not-p* brings it about that one no longer knows that *p*. Lewis is referring specifically to the knowledge involved in presupposing and ignoring, as when, in telling the time when glancing up at the church clock, you presuppose that the clock is in good working order. Such knowledge is knowledge, Lewis holds,

> but it is an *especially* elusive sort of knowledge, and consequently it is an unclaimable sort of knowledge. You do not even have to practice epistemology to make it vanish. Simply *mentioning* any particular case of this knowledge, aloud or even in silent thought, is a way to attend to the hitherto ignored possibility, and thereby render it no longer ignored, and thereby create a context in which it is no longer true to ascribe the knowledge in question to yourself or others. (1999, p. 438)

All knowledge that is not completely certain *is* knowledge, for Lewis, but not claimable knowledge. And so in the Mont Blanc case: the referents of our everyday terms will in some contexts include certain objects as parts whenever those are, in those contexts, not projected as distinct. Yet this sort of parthood is elusive: it is never claimable, since to claim it would amount to a shift in context.

Consider your thirsty brother's utterance: 'This glass is empty', in the context of a drinking session in your local pub. Suppose that there are, as a matter of fact, tiny drops of beer at the bottom of the glass. Your brother's partition of reality does not recognize these drops of beer. His utterance is, in the given context, true. But the sentence 'this glass is empty' entails the sentence:

> [B] There are no tiny drops of beer in this glass.

Since the latter is false, the mentioned entailment must thus be unavailable to your brother in that context, since it is a context in which 'this glass is empty' is both judgeable and true. That it *is* unavailable can now be understood as follows: both [B] and its negation would undermine a context of the sort your brother is currently inhabiting. Moving from 'this glass is empty' to 'this glass contains tiny drops of beer' amounts to moving to a *more refined context*. In this more refined context, the utterance 'this glass is empty' is false.

5. Fred's House

Let us now turn to the already mentioned problem of Fred's house. On Lewis's non-contextualized solution to the problem of the many, the sentence

> [C] The garage is not a part of Fred's house

falls into a supertruth-value gap. [C] is true on some precisifications and false on others because, as Lewis (1993, p. 172) puts it, 'no established convention or secret fact' decides the issue of whether or not the garage is part of Fred's house.

When context is taken into consideration, however, the need to acknowledge supertruth-value gaps falls away. We still need to recognize three different alternatives as far as the corresponding *sentences* are concerned. Now, however, these have the labels: *judgeable and true*, *judgeable and false*, and *not judgeable*.

An utterance P is supertrue if and only if:

> (T1) the utterance successfully imposes in its context C a partition of reality in which portions of reality corresponding to its singular terms are foregrounded, *and*
> (T2) the corresponding families of determinate aggregates consistent with this foregrounding are such that, however we select corresponding aggregates, P is true.

'Bruno is in the living room' is true when 'Bruno' singles out the cat aggregates you are attending to (T1) and all of those aggregates are in fact in the living room (T2).

An utterance P is superfalse if and only if *either*:

> (F0) it fails to impose in its context C a partition of reality in which portions of reality corresponding to its singular terms are foregrounded, *or both*:
> (F1) the utterance successfully imposes in its context C a partition of reality in which portions of reality corresponding to its singular terms are foregrounded, *and*:
> (F2) the corresponding families of determinate aggregates consistent with this foregrounding are such that, however we select corresponding aggregates, P is false.

Suppose Bruno is in the kitchen, but your bleary-eyed husband, looking at a cat-shaped piece of furniture in your living room, utters: 'Your cat is in the living room'. This utterance is then superfalse in virtue of (F0). There are no feline portions of reality to sustain a partition of the needed sort. Suppose that you look at Bruno in the kitchen and utter 'Bruno is a unicorn'. This utterance is superfalse because (F1) and (F2) are satisfied. Your utterance does single out the family of aggregates that constitutes Bruno, but it is false of every single one of those aggregates that it is a unicorn.

To see how this works for Fred's house, suppose you are approached by a stranger and you assert [C]. You dimly remember the plans of Fred's house as including a boundary dividing garage from house. Your utterance, accordingly, rests on a conception of garage and house as discrete items: it would thus impose upon the reality within the neighborhood of Fred's house a boundary of the appropriate sort. If the conditions as imposed by Lewis are satisfied, however, then there is no such boundary in reality. Hence your attempt to impose a partition of the given sort fails, and your utterance is superfalse (by F0).

Fred's house is fictional. Consider the following real case. Switzerland, Germany, and Austria meet in the heart of Europe somewhere in the neighborhood of Lake Constance. No international treaty exists which establishes where, in or around Lake Constance, their respective borders lie. This still occasionally gives rise to disputes, for example as concerns the rights to fish in different portions of the Lake. Suppose, now, that

you point to a certain kilometer-wide volume of water in the center of the Lake, and you assert:

[D] That water is in Switzerland.

Here, too, there is no established convention or secret fact which decides the issue. What this means, however, is not that [D] asserts a truth on some evaluations and a falsehood on others. Again [D] is simply (super)false, by (F0). Whoever uses [D] to make an utterance in the context of current international law is making the same sort of mistake as is your bleary-eyed husband who looks at some cat-shaped piece of furniture in the living room and judge that your cat Bruno is in your living room (even though Bruno is in fact in the kitchen). In both cases, reality is not such as to sustain a partition of reality of the appropriate sort.

Consider, finally, the 'Buy one. Get one free' case. When you come home with two packs of Marlboro, you look at one of the packs and you say:

[E] This is the one I bought.

[E] does not exemplify a supertruth-value gap. For either the merchant ran one of the packs through his scanner to get the price. You then ended up with two packs, but you only paid for one, and in this case there is no supertruth-value gap. Or the merchant put the two packs together under his scanner, and you paid half price for both. A judgment to the effect that you bought one and got one free is then just superfalse by (F0). Suppose instead that the scanner is programmed to calculate the price for cigarettes only *after* all items have been scanned – it needs to see, first, how many packs you buy. In that case 'Buy one. Get one free' is false advertising. You do not get one free when you buy the other. Rather, again, you get two packs for the price of one.

This, now, tells us more clearly what the friend of supertruth-value gaps needs to find. Such gaps can arise only if (T1) (and thus also (F1)) is satisfied. But it remains to be determined whether there are, in fact, any naturally occurring utterances (judgeable in natural contexts) that are such as to satisfy this condition.

6. Moral Contexts

The above solution to the problem of vagueness in terms of contextualized supervaluation can be useful in solving problems of vagueness in moral philosophy also. Consider the following intuitive metaethical principle, sometimes called 'the Principle of Access': *If an act is obligatory, then its agent can know that it is obligatory*. If this principle is true, then there are no obligations unknowable to their bearers. This principle has much intuitive appeal because it seems that, if we are to be able to blame people for not doing their duty, then obligations must be knowable. An agent can *choose* to do her duty based on the right motives only if she can know her obligations. And only then, when she can *choose* to do her duty, can we also correctly blame her when she fails to do her duty. If an agent's excuse for not doing her duty is that she did not know whether or not she had an obligation, then we can say that she *ought to* have known better. But we cannot correctly say that she ought to have known better if the obligation in question is itself unknowable.

Sorensen (1995) has argued against the Principle of Access. One of his main arguments against this principle is that there are vague terms in our language. For example, if Scrooge promises to pay a fair wage to his clerk, what is the minimum he can pay and still keep his promise? If 'fair wage' is a vague term, then it would normally be unknowable in principle what exactly a fair wage is. Similarly, nobody could ever know exactly what a minimum fair wage is, since there is no one wage that deserves to be called '*the* minimum fair wage'. If the agent cannot know what the minimum fair wage is, then it seems that he cannot know either that he is obliged to pay this amount.

Notice that Scrooge can know *that* he is obliged to pay the minimum fair wage without knowing what the minimum fair wage is, just as he can know that the tallest spy is involved in espionage without knowing who the tallest spy is. He can, in other words, know *de dicto* that he is obliged to pay his clerk the minimum fair wage. But if it were mere *de dicto* knowledge which were at stake here, then Sorensen's objection would not affect the Principle of Access. The Principle of Access must be referring to *de re* knowledge in order for Sorensen's objection to have any force at all. If the minimum fair wage is unknowable, then Scrooge cannot know *de re* that he is obliged to pay the minimum fair wage: he cannot know that he is obliged to pay this-or-that amount of money. And if he cannot know that he is obliged to pay this-or-that amount of money, then he is not obliged to pay the minimum wage. Yet it seems to be a fact that people in Scrooge's situation *are* obliged to realize their promises even if they make their promises by means of language that contains vague terms, and this I think is Sorensen's real worry.

But this worry is misplaced. We do not have to throw the Principle of Access overboard. We do not have to accept that there are unknowable obligations. And here the contextualized version of supervaluation comes to our rescue. Scrooge does indeed have the knowledge required to make a promise, and he *does* indeed make a promise that can and will be carried out. For in the context that is pertinent to his promise, decisions have already been made concerning what the minimum fair wage is. It is not the case that any old minimum fair wage would be just as good as any other. Rather, a certain wage is – *in the context of a given society with its minimum-wage laws* – the minimum fair wage.[6] We do not have to throw away the Principle of Access on the ground that there are vague terms in our language. Scrooge can go and find out what the minimum fair wage is in his society and he can thereby fulfil his promise. His obligation is not unknowable, even if it turns out that he himself happens not to know what the minimum fair wage is at the present time.

But suppose now that you promise to paint your brother in law Fred's house while Fred and your sister are on vacation. As you arrive at Fred's house with your paint and

6. When we have contexts in which people make fiat decisions A *is* B just in case A *counts* as B (see Searle 1995). For example, if people make a decision about drawing a certain line on a map and if this line *counts* on all sides as the border of a country, then it *is* the border of that country. We can of course suppose that Scrooge lives in a society in which no such fiat determination of the concept 'minimum fair wage' has been made. Then, indeed, it is true that Scrooge can not use a sentence like 'I promise to pay you the minimum fair wage' in order to effect a genuine promise.

your brush it strikes you that you do not know whether 'Fred's house' refers to the house with or the house without the (very large) garage. Indeed when you look inside you discover that the garage, too, is inhabited – there are beds and a kitchen inside it, so that you are not now sure whether it might not itself be what was intended by the phrase 'Fred's house'. You have promised to paint Fred's house, so it seems that you are obliged to paint his house. But according to the Principle of Access, you are obliged to paint his house only if you can know that you are obliged to paint his house. And certainly it would not be good enough merely to have the knowledge *de dicto* that you are obliged to paint his house. You must have the knowledge *de re* that you are obliged to paint this or that building or complex of buildings.

Perhaps you remember the plans of Fred's house as including a boundary dividing garage from house. Then there is no problem. It is simply superfalse that the garage is a part of Fred's house. Since it is superfalse that the garage is a part of Fred's house, your promise is to paint the house without the garage. You can thus fulfil your promise by painting the house but leave the attached garage as it is.

But perhaps there is no such boundary dividing the garage from the house. In that case, I shall argue, you do not succeed in making a promise by uttering 'I promise you to paint your house'. The reason is as follows. An agent might have an obligation even if the obligation cannot be carried out. Denying that agents can have obligations even when they cannot be carried out amounts to accepting the principle that an act is obligatory only if the agent can carry out the obligation. We need not accept this principle. Promise making, however, is a special way of bringing an obligation into existence. It is reasonable to demand that a speech act, in order to be a promise, must give rise to an obligation which can be carried out. For when making a promise the agent is making a direct and conscious choice to bring an obligation into existence. Other obligations are not directly and consciously chosen by the agent. For example, an obligation may come into existence because a general moral rule (whatever its ontological status) is applicable in the given circumstances. The general moral rule which says that we should speak the truth is applicable in those circumstances in which we are speaking. Whenever we speak we are obliged to speak the truth. In such cases, however, although the agent chooses to speak or not, he is not directly choosing to bring the obligation to speak the truth into existence. Rather, he is merely choosing to bring himself in the circumstances in which the given moral rule happens to apply. Given the special way in which promises bring obligations into existence, we can reasonably require that an agent should be able to fulfil his promises without requiring that all obligations are such that they can be carried out.

Given the above demands on promises, consider the sentence 'the house without the garage is Fred's house'. This sentence is false given that there is no boundary dividing the house from the garage in the context in which you were making your promise. Similarly, the sentence 'the garage is Fred's house' is false. If you had used the sentence 'I promise to paint your house' to make a promise, then you should have been able to fulfil your promise. You have fulfilled your promise just in case you paint Fred's house. But no matter which building you paint, it will be false that that building is Fred's house, and so you will have failed to paint Fred's house. There is, in other words, no way in which you can carry out your promise. Hence, you failed to use the sentence 'I promise to paint your house' to make a promise. (This is a failure parallel to that labeled (F0) in the above.)

Fred might be disappointed when he comes home because he expects to find that his house has been painted, but you did not fail to fulfil your obligation, for you had in fact no obligation. Next time, Fred will have to remember to tell you whether he wants you to paint the house with or without the garage.

There is thus no need to throw away the Principle of Access on the grounds that there are vague terms in our language. There *are* indeed vague terms in our language, and people do indeed make promises. But the method of contextualized supervaluation shows us that this vagueness is rendered innocuous when we take the relevant truth-bearers to be, not sentences, but judgments in naturally occurring contexts.[7]

References

Fine, K. 1975. Vagueness, Truth and Logic, *Synthese* 30, 265–300.

Lewis, D. 1996. Elusive Knowledge, *Australasian Journal of Philosophy* 74, 549–567.

Lewis, D. 1993. Many, But Almost One, in J. Bacon, K. Campbell & L. Reinhardt, *Ontology, Causality and Mind: Essays in Honour of D. M. Armstrong*, Cambridge: Cambridge University Press, 1993. Cited as reprinted in *Papers in Metaphysics and Epistemology*, Cambridge: Cambridge University Press, 1999, 164–182.

Morreau, M. 1999. Supervaluation Can Leave Truth-value Gaps after All, *Journal of Philosophy* XCVI, 3, 148–156.

Searle, J. R. 1983. *Intentionality. An Essay in the Philosophy of Mind*, New York: Cambridge University Press.

Searle, J. R. 1995. *The Construction of Social Reality*, New York: The Free Press.

Sorensen, R. 1995. Unknowable Obligations, *Utilitas* 7/2, 247–271.

Sorensen, R. 2000. Truthmaker Gaps and The No-No Paradox, manuscript.

Smith, B. and Brogaard, B. 2001. A Unified Theory of Truth and Reference, forthcoming in *Logique et Analyse*.

Unger, P. 1980. The Problem of the Many, *Midwest Studies in Philosophy* 5, 411–67.

7. Thanks are due to the National Science Foundation which supported work on this paper under Research Grant BCS–9975557: "Geographic Categories: An Ontological Investigation". I would also like to thank Philipp Keller, Pierluigi Miraglia, Kevin Mulligan, Joe Salerno, John Searle, Frederik Stjernfelt, Joe Tougas, and Achille Varzi for helpful comments at the presentation of this paper at the 23rd Wittgenstein Symposium. Thanks are due also to David Suits for numerous discussions about vagueness.

Sind alle Religionen gleichermaßen rational?

ANDRZEJ BRONK

A. Einführende Bemerkungen.

1. Zielsetzung. Auf die Titelfrage, *Sind alle Religionen gleichermaßen rational?*, gibt es keine allgemeine, eindeutige Antwort und – was noch wichtiger ist – es kann auch keine geben. Zwar werden Leute in der westlichen Hemisphäre sehr wahrscheinlich intuitiv mit einem „Nein" reagieren, aber eine Begründung dieser Antwort wird ihnen schon Schwierigkeiten bereiten. Zunächst darum, weil die Antwort von einer früheren Antwort auf andere Fragen abhängig ist, und weil die betreffenden Begriffe – von Religion und von Rationalität – vieldeutig sind. Als Erstes muß also festgestellt werden, was unter Rationalität von Religion verstanden wird und um welche Aspekte von Religion es geht. Je nach Präzisierung der Begriffe von Rationalität und von Religion – und es gibt unter Epistemologen und Religionswissenschaftlern diesbezüglich keine Übereinstimmung – verfügt man über eine breite Palette von Antworten.

Trotz des Titels meines Referates geht es mir aber im Folgenden nicht darum, Kriterien und Standards der Rationalität von Religion zu entwerfen, um dann einzelne Religionen unter dem Gesichtspunkt der Rationalität zu bewerten und eventuell eine Hierarchie der Religionen in Hinsicht auf ihre größere oder geringere Rationalität zu erstellen und vielleicht eine Antwort zu geben, welche von den vielen Religionen mehr oder weniger rational ist. Wenn schon die Religionen verglichen wurden, dann unter dem Gesichtspunkt ihrer Wahrheit. Ich verstehe mein Ziel kantisch als eine Aufgabe, die die apriorischen Bedingungen zur Frage der Rationalität (Rationalitätsbedingungen) von Religion ans Licht zu bringen hat, und als Prolegomena zu einer Analyse des Begriffs der Rechtfertigung religiöser Sätze. Ich nütze die Frage, ob alle Religionen gleichermaßen rational sind, als Ausgangspunkt und Gelegenheit, um einige generelle Probleme der Anwendung des Begriffs der Rationalität auf die Religion zu diskutieren. Ich konzentriere mich dabei auf den kognitiven (doktrinären) Aspekt von Religion – auf den Glauben und auf die religiöse Lehre – und übergehe solche Aspekte der Religion, wie etwa den praktischen, den kultischen oder den institutionellen Aspekt.

Die Frage nach dem Begriff der Rationalität von Religion ist zunächst eine theoretische Frage, zu deren Lösung keine detaillierten empirischen Untersuchungen über den faktischen Glauben und die religiösen Praktiken konkreter Gläubiger durchgeführt werden müssen. Obwohl ich mich aber für den Begriff der Rationalität von Religion interessiere, gilt meine Aufmerksamkeit der Sache selbst, d.h. der Frage, worin die Rationalität von Religion besteht – oder eher – bestehen kann. Ich bin weit davon entfernt, Grundzüge einer allgemeinen Theorie der Rationalität von Religion, ähnlich dem Modell der wissenschaftlichen Rationalität (I. Lakatos), projektieren zu wollen, das sich auf alle Religionen anwenden ließe, weil mir das heute aus mehreren Gründen nicht realisierbar erscheint. Es geht mir auch nicht darum, Regeln aufzustellen, wann sich Leute rational benehmen,

B. Brogaard, B. Smith (Hg.), *Rationality and Irrationality / Rationalität und Irrationalität*, S. 46–61.

wenn sie sich für Religion entscheiden, weil das historische, soziologische und psychologische Untersuchungen verlangen würde. Außerdem, lässt sich nicht immer eindeutig feststellen, woran ein religiöser Mensch wirklich glaubt und wie er seine religiösen Meinungen (seinen Glauben) und seine religiöse Handlungen (Praktiken) rechtfertigt. Desto weniger geht es mir um eine Apologie von Religion im Allgemeinen oder einer bestimmten Religion, z.B. des Christentums, unter dem Aspekt ihrer Rationalität. Ich versuche den Begriff der Rationalität von Religion im weiteren Sinne zu verstehen, und nicht, wie das oft der Fall ist, auf das Christentum allein zu beziehen. Es bleibt dennoch ein paradigmatischer Fall von Religion. Konkret begrenze ich mein Interesse u.a. auf folgende Fragen: In welchem Sinne kann von der (epistemischen) Rationalität von Religion im allgemeinen oder einer bestimmten Religion (z.B. des Christentums) gesprochen werden? Kann der allgemeine Begriff der Rationalität im Bereich der Religion – und wenn ja, in welchem Sinne – angewendet werden? Welche Faktoren und Gründe müssen berücksichtigt werden, um eine Religion als rational (oder irrational) zu bezeichnen?

2. Aktualität der Frage. Die Bedeutung, die dem Problem der Rationalität von Religion zugeschrieben wird, ist charakteristisch für die abendländische Kultur und anderen, außereuropäischen Kulturen in diesem Maße unbekannt. Sie erlebten nie in diesem Maße die Spannung zwischen dem Glauben und der Vernunft und zwischen der Religion und der Wissenschaft.

Die Frage nach der Rationalität von Religion hat zuerst die Philosophen und die Theologen, weniger die Religionswissenschaftler interessiert. Heute, durch das Bewusstsein von Pluralität und Verschiedenheit der Religionen, bekommt sie eine neue theoretische und praktische Bedeutung. Der Mensch in der abendländischen Kultur möchte ungern eines irrationalen Denkens oder einer irrationalen Handlung beschuldigt werden. Auch der Gläubige, wie andere Menschen, möchte rational denken und handeln und für ihn ist es schon von Bedeutung, ob das, woran er glaubt, wahr (rational) sein könnte, oder nur eine subjektive Illusion ist, u.a. darum, weil diese Entscheidung einen gravierenden Einfluß auf sein Denken und somit auf sein Verhalten ausübt. Wird man ihn fragen, ob Religion, besonders seine eigene, rational ist, wird er höchst wahrscheinlich spontan „Ja" sagen, obwohl er vielleicht seine Antwort nicht näher erklären und begründen können wird. Diese positive Einstellung zu der Rationalität von Religion kommt unter anderem davon, dass der Begriff „Rationalität" in der europäischen Kultur generell ein axiologisch positiver und normativer Begriff ist. „»Unvernünftig« impliziert für jedermann einen Vorwurf" (L. Wittgenstein 1971, 93).

In einer Zeit der Säkularisierung der Kulturen und der Vielheit und Verschiedenheit der in der Welt auftretenden Religionen kann ein religiöser Mensch nicht umhin zu fragen, ob sie alle gleich wahr und rational sind und ob die Wahl von einer sich auf rationale Gründe stützen kann. Es bestehen gravierende Differenzen und Widersprüche zwischen den Religionen in der Lehre (über Gott, die Welt, den Menschen), in der Moral, im Kult, in institutioneller Organisation. Jede Religion beruft sich auf ihren Anfang in eigener Offenbarung und auf eine eigene Stifterfigur. Selbst das Christentum divergiert stark in den Aussagen über z.B. die Erbsünde, den freien Willen, das Leben nach dem Tode usw. Man kann aber anderseits sehen, wie Leute, abgestoßen vom einseitigen Rationalismus eigener Religion und auf der Suche nach einer gefühlsmäßigen Heimat, aus den überrationali-

sierten Kirchen in die religiösen, mehr Geborgenheit versprechenden Minderheiten (Sekten) fliehen. Im Christentum hat sich die Lage insoweit geändert, dass für viele Christen das doktrinäre (rationale) Element nicht mehr so wesentlich ist wie früher, oder wie es die offizielle, institutionelle Kirche versteht.

3. Geschichte der Diskussion. Seit Jahrhunderten ist das Problem der Rationalität von Religion ein wohl bekanntes, obwohl nicht immer unter diesem Namen diskutiertes Problem. Ziemlich oft haben die Philosophen mit Hilfe der Vernunft und des Begriffs der Rationalität die Religion sowohl unterstützt als auch kritisiert. Schon Xenophanes von Kolophon verspottete die antropomorphen Gottesvorstellungen seiner Zeitgenossen. Einen religiösen Agnostizismus vertrat Protagoras, als er sagte, dass es keine Möglichkeit gebe, etwas über die Götter zu wissen, weder, ob sie seien, noch, ob sie nicht seien. Im Mittelalter wurde die Frage nach der Rationalität von Religion als das Problem des Verhältnisses von Religion und Philosophie, von Glaube und Vernunft (*fides et ratio*), dann von Religion und Wissenschaft gestellt. Klassische Lösungen finden sich bei Averroës, Anselm von Canterbury und Thomas von Aquin.

Eine besondere Aktualität gewann das Problem der Rationalität von Religion seit der Aufklärung. Die Überzeugung, dass der Gebrauch der Vernunft für Religion unerlässlich sei, war ein Allgemeingut unter den damaligen Philosophen (und Theologen). Es war aber auch die Aufklärung, die den modernen Begriff der „Religion als Unvernunft" prägte. Die Rationalität eines Satzes mit seiner Wahrheit, in einem scharfen Gegensatz zur Vernunft, gleichsetzend, erklärte sie alle historischen Religionen als Aberglauben und Ort purer falscher Sätze, und warf den Anhängern einer Offenbarungsreligion (des Christentums) vor, dass sie sich irrational, d.h. im Widerspruch zur Vernunft verhalten. B. Pascal, der das Christentum verteidigen wollte, meinte, es sei aus pragmatischen Gründen (Pascalische Wette) »vernünftig«, sich für die Existenz Gottes und den (christlichen) Glauben zu entscheiden. J. Locke schrieb *The Reasonableness of Christianity: As Delivered in the Scriptures* (1695), wo er u.a. die These vertrat, dass das, was Gott geoffenbart habe, zwar unbedingt wahr sei, was aber als göttliche Offenbarung akzeptiert werden könne, darüber entscheide nur die Vernunft.

In der Neuzeit waren es vor allem K. Marx und S. Freud, die der Religion die Irrationalität vorwarfen und sie als Projektion menschlicher Wünsche, Sehnsüchte und Ängste sahen, während E. Durkheim und M. Weber die (christliche) Religion für eine Verteidigerin der Rationalität hielten. Am Ende des XIX. Jh. fühlte sich jeder aufgeklärte Mensch (Wissenschaftler) gezwungen, eigene Religiosität zu leugnen, weil Religion im Vergleich zur Wissenschaft völlig irrational oder zu wenig rational erschien. Im XX. Jh. trat – im Namen der Wissenschaft und Logik – der logische Positivismus (R. Carnap, J. L. Austin, A. J. Ayer) mit einer radikalen Kritik der Religion auf: Er sprach Sätzen, die religiöse Inhalte betrafen, jeden epistemischen Wert ab, da solche Sätze weder verifizierbar, noch falsifizierbar sind; daher seien sie un-sinnig, bzw. Sinn-los.

4. Das dialektische Gesicht der Rationalität. In einer Diskussion um die Rationalität von Religion ist es schwer, einen neutralen Standpunkt zu bewahren: sie war und ist immer mehr oder weniger ideologisch geprägt. Hauptsächlich ist es ein Problem der Theisten. Für die Atheisten, die für sich die Tradition des Rationalismus in Anspruch nehmen und

sich oft als die einzigen authentischen Verteidiger der Vernunft darstellen, ist mit der Negierung der theistischen These: »Gott existiert«, die Sache der Religion als etwas Irrationales eigentlich schon erledigt.

Bemerkenswert ist die Divergenz, mit welcher die Theisten und die Atheisten den Begriff der Rationalität (Irrationalität) von Religion verstehen. Mann kann sich schwer einen Theisten vorstellen, der die Rationalität von Religion angreifen, und einen Atheisten, der sie verteidigen würde. Die Theisten berufen sich auf die Vernunft und die Logik, um die Existenz Gottes zu beweisen, während gleichzeitig die Atheisten mit derselben(?) Vernunft und Logik das Gegenteil beweisen. Mit dem Beweis der Rationalität von Religion wollen die Theisten die Existenz letzterer rechtfertigen, die Atheisten demgegenüber sie mit dem Nachweis ihrer Irrationalität als unwahr zurückzuweisen. Ein rational denkender Theist, überzeugt, dass die Vernunft (der Intellectus) ein beliebtes Geschöpf Gottes ist (Nikolaus von Kues), ja eine konstitutive Eigenschaft Gottes sein müsse, verteidigt die Religion, weil sie rational ist, ein rationalisierender Atheist sieht Religion – auch im Namen der Vernunft – als eine Sphäre des Irrationalen. Die Situation, dass die Theisten spontan unter allen Umständen zu beweisen versuchen, dass Religion die allgemeinen Kriterien von Rationalität erfüllt, und dass auf der anderen Seite die Atheisten genau so entschlossen diese These bestreiten, gibt zu denken. Ist es wirklich dieselbe Vernunft und Logik, die Thomas von Aquin die These von der Existenz Gottes anerkennen und Carnap sie bestreiten lässt? „Wenn ein Atheist sagt »Es gibt kein Jüngstes Gericht«, und jemand anders sagt »Es gibt«, meinen sie dasselbe?" (L. Wittgenstein 1971, 94).

B. Kontext der Diskussion.

1. Rationalität und Wissenschaft. Bis in die jüngste Zeit sah man in der Rationalität einen unerlässlichen Bestandteil der abendländischen Kultur und hielt sie für einen autonomen und – logisch, epistemisch, auch ethisch – hochgeachteten Wert. War es im Altertum und Mittelalter die Philosophie, die als Hintergrund der abendländischen Diskussion über die epistemische Rationalität von Religion diente, sah man seit der Neuzeit in der Wissenschaft den Inbegriff aller Rationalität. Eine szientistische Einstellung, die die Rationalität mit der Methode, der Logik, der Widerspruchsfreiheit und der Begründung identifizierte, akzeptierte allein den Begriff der Rationalität und die Begründungsverfahren der exakten Wissenschaften als Standards von Rationalität.

Von diesem Standpunkt aus und gemessen an der Rationalität der Wissenschaft musste jede Religion als Domäne des Irrationalen erscheinen. Keine Religion ist rational in diesem Sinne, dass sich alle ihre Glaubenswahrheiten empirisch verifizieren oder falsifizieren lassen. Verhältnismäßig einfach lassen sich in jeder Religion Sätze finden, die mit den Thesen der Wissenschaft unvereinbar sind. Die Naturwissenschaften zeigen, dass die Welt ohne Gott (*God of gaps*) auf einem natürlichen Wege erklärt werden kann. In einer Situation, in der das Ideal der wissenschaftlichen Rationalität dominiert und der Bereich der Tatsachen und der Logik zur Wissenschaft zugehört, bleibt natürlich wenig Platz für Religion, für Über-natürliches und Transzendentes. Einer Rationalisierung der Welt durch Wissenschaft entspricht eine Derationalisierung von Religion und in der Konsequenz auch eine Säkularisierung. Ist es aber Aufgabe der (empirischen) Wissenschaften

oder sogar der Philosophie eine Art Basis für die Religion darzustellen?

2. Gegenwärtiger Kontext der Diskussion. Heute – im Vergleich zum XIX. Jh. – verläuft die Diskussion über das Problem der Rationalität von Religion in einem kulturell und wissenschaftlich geänderten Kontext. Die neue epistemische Situation besteht u.a. darin, dass die Rationalität – des Menschen, der Welt und der Wissenschaft – grundsätzlich in Frage gestellt wurde. Kein anderer Begriff ist in der letzten Zeit so oft und heftig diskutiert und kritisiert – seltener verteidigt – worden als der der Rationalität. Merkwürdigerweise teilt den Zweifel an der Rationalität heute auch das atheistische Denken, das sich von alters her als DER Fürsprecher und Verteidiger der Rationalität verstand.

Der Mensch von der Straße akzeptiert nicht mehr spontan die Tatsache seiner eigenen Rationalität und steht der Vernunft und der Wissenschaft mit einem mehr oder weniger tiefen Misstrauen gegenüber, und die Vorstellung von der Rationalität des Menschen erscheint ihm wie ein Aberglaube (I. M. Bocheński). Auch unter den Philosophen ist es keine Selbstverständlichkeit mehr, dass die Rationalität ein fundamentaler philosophischer Begriff und eine Qualifikation (ein Wert) von entscheidender Bedeutung ist, die den Meinungen als epistemische (theoretische) und den Handlungen als pragmatische (praktische) Rationalität zukommt. Ebenfalls wird die Welt heute radikal anders gesehen als z.B. im Altertum oder im Mittelalter: sie wird nicht mehr als ein harmonischer, intelligibler Kosmos aufgefaßt, sondern als eine zufällige, chaotische Anhäufung von Objekten. Nach Th. Kuhn beweist die Wissenschaftsgeschichte nicht den Anspruch der Wissenschaften auf eine absolute, außerparadigmatische Rationalität. Sie zeigt eher, dass es keine festen, expliziten, außerhalb jeder Tradition (des Paradigmas) stehenden Regeln des wissenschaftlichen Vorgehens gibt. Nachdem die Wissenschaft ihre Anschaulichkeit, Klarheit und Gewissheit verloren hat, erscheint manchen auch die Religion weniger irrational.

3. Christentum und die Rationalität von Religion. Die Dominanz des Christentums im Abendland machte es aus, dass die Frage nach der Rationalität von Religion – zu Unrecht – oft auf die Frage nach der Rationalität des Christentums reduziert wurde, das als ein theoretisches Begriffssystem mit der zentralen These »Gott existiert« verstanden wurde. Es ist aber wahr, dass das Christentum von Anfang an viel Aufmerksamkeit und Mühe der Anpassung seiner Glaubenswahrheiten an die Vernunft (Philosophie und Wissenschaft) gewidmet hat und in seiner Sorge um eigene Rationalität eine eigenartige Stellung unter den Religionen einnimmt. Die offizielle Lehre der römisch-katholischen Kirche (*Magisterium*) widersetzte sich immer einer einfachen Gegenüberstellung des Glaubens und der Vernunft, sowie der These, dass sich Glaube und Vernunft im Widerspruch befänden. Glaube braucht Vernunft und der Mensch soll auf vernünftige Weise glauben. Sie verficht die These von der Rationalität Gottes (obwohl nicht notwendig nach den Regeln menschlicher Rationalität) und der Welt, setzt sich für die Rationalität der Glaubenswahrheiten und der Religion ein und legt Wert darauf, die christlichen Glaubenswahrheiten mit den Ergebnissen der Wissenschaft in Einklang zu bringen. Sie untermauert die Rationalität der Welt und des Menschen (als Ebenbild Gottes) mit der religiösen Idee, dass sie Geschöpfe Gottes, der höchsten Intelligenz, sind. Der doktrinäre Anteil (*Credo*) wird im Christentum besonders stark betont und die Gläubigen, die für die christliche

Glaubenswahrheiten sterben, werden als Märtyrer geehrt.

Die Tatsache, dass sich Gott dem Menschen offenbarte, bedeutet nicht, dass er ihm das Recht auf die Vernunft absprach. Die römisch-katholische Kirche verurteilt den Verzicht auf die Vernunft in Glaubenssachen als Fideismus. Andererseits lehrt sie, dass der Glaube eine Gottesgnade ist. Das Vaticanum I verurteilte die Behauptung, dass Gott als Schöpfer der Welt nicht mit Hilfe der natürlichen Vernunft erkannt werden könne. Ein neuestes Zeugnis für die Bemühung der katholischen Kirche um ihre eigene Rationalität ist die Enzyklika *Fides et Ratio* Johannes Paul II. (1998). Das Interesse des Christentums an der Rationalisierung der Lehre findet seinen Ausdruck in der Ausbildung einer christlichen Theologie (Dogmatik), die kein Äquivalent in anderen großen Religionen hat, was ihm manchmal den Einwand der Überrationalisierung der Religion bringt. In diesem Kontext vertritt Karl Barth unter den Theologen die These, dass die Absicht, der Religion die Rationalität zuzusprechen, einer Verkennung ihrer wirklichen Natur gleich komme.

Wie die vielen spekulativen theologischen Systeme der Vergangenheit und der Gegenwart zeigen, ist der Wunsch nach einer einheitlichen Theologie nie in Erfüllung gegangen und vielen Katholiken ist die Mehrzahl der theologischen Entwürfe kaum bekannt. Dabei gibt es auch im Rahmen des Christentums Strömungen, die den Glauben radikal der Vernunft entgegen stellen. Der Vernunft feindliche Töne lassen sich z.B. im ersten Brief Paulus an die Korinther hören: „Hat Gott nicht die Weisheit der Welt as Torheit erwiesen?" (1 Kor 1, 20). „Jeder, der die Briefe des Apostels Paulus liest, findet es ausgesprochen: nicht nur, dass der Glaube nicht vernünftig ist, sondern daß er eine Torheit ist" (L. Wittgenstein 1971, 93). Im christlichen Altertum hat Tertullian (+222) nicht nur behauptet, dass nur das Christentum im Besitz der vollen Wahrheit sei, sondern die These aufgestellt, dass, wenn das Christentum auch das Unbegreifliche verkünde, es geglaubt werden müsse: *certum est, quia impossibile est, quia absurdum est*. Der Islam wirft dem Christentum Irrationalität vor, indem er es für eine polytheistische Religion hält, die mit dem Trinitätsdogma die Existenz „dreier Götter" hinnimmt.

C. Religion und Rationalität.

1. Begriff der Religion. Die Diskussionen über die Definition und den Begriff (Umfang und Inhalt) der Religion, nehmen kein Ende. Es gibt viele Typen von Religionsdefinitionen, oppositionelle oder komplementäre, die einen bestimmten Aspekt einer Religion beschreiben, und keine von ihnen wird allgemein akzeptiert. Oftmals, wenn in der abendländischen Kultur von Religion die Rede ist, wird vor allem das Christentum gemeint. Es gibt aber andere Weltreligionen, wie das Judentum, den Islam, den Hinduismus, den Buddhismus, den Konfuzianismus, den Taoismus, von den hunderten und tausenden kleineren und größeren Religionen in der Vergangenheit und Gegenwart nicht zu sprechen, die auch in die Definition von Religion mit einbezogen werden müssen.

Eine institutionelle Religion lässt sich multidimensional von ihrer doktrinären, kultischen und sozial-institutionellen Seite beschreiben. Sie umfasst einen Komplex von Glaubenssätzen, religiösen Meinungen, Symbolen, Normen, Lebensformen, Handlungen, Haltungen, Praktiken und Institutionen. Das, was allen Religion gemeinsam ist, wird sehr vage als Glaube an Gott oder eine höhere Macht genannt. Wie kann man aber fest-

stellen, ob alle Religionen wirklich an denselben Gott glauben? Es gibt den aufklärerischen Begriff der natürlichen Religion (*religio naturalis et rationalis*) in der Gegenüberstellung zu allen historischen Offenbarungsreligionen. Er ist ein klarer Ausdruck der rationalisierenden Tendenz der Philosophen mit Hilfe der glaubensunabhängigen natürlichen Theologie ein Wissen über den »Gott der Philosophen« zu erlangen. Diese Probleme – unter dem Motto: „es gibt keine Religion, sondern nur Religionen" – haben manche Wissenschaftler zu einem Verzicht auf eine generelle Definition von Religion gebracht.

2. Begriff der Rationalität. Obwohl der Begriff der Rationalität im Zentrum vieler philosophischen und wissenschaftstheoretischen Diskussionen steht, gibt es keine Einigung darüber, wie er zu verstehen ist. Analysiert werden die Natur der Rationalität, ihre Definition, Träger, Kriterien, Gründe, Ziele und Typen. Als einem axiologischen Begriff wird ihm grösstenteils ein positiver Wert beigemessen. Die These, dass rationales Denken ein autonomer und höchster Wert ist, schließt aber nicht die Möglichkeit aus, dass es auch andere Wege zur Erkenntnis, z.B. einen mystischen, gibt, oder dass die Emotionen als ein unerlässlicher motivierender Faktor der menschlichen Handlungen zu akzeptieren sind. Es gibt auch Epistemologen (W. W. Bartley) nach denen die Parteinahme (*commitment*) für epistemische Rationalität immer eine irrationale, d.h. rational nicht zu rechtfertigende Entscheidung ist. Andere, wie L. Kołakowski, behaupten, dass sowohl die Art der Beschäftigung mit der Epistemologie wie auch die Bestimmung des Begriffs der Rationalität die Anerkennung der Existenz Gottes voraussetzt. Es gibt aber keine guten Argumente für die These, dass die Existenz Gottes eine Vorbedingung für die Gesetze der Logik und somit für ein rationales Denken sei.

Als Typen der Rationalität werden genannt: die metaphysische (substanziale) Rationalität des Seins (der Welt), die funktionale Rationalität der Meinungen (epistemische, theoretische Rationalität) und der Handlungen (pragmatische, praktische Rationalität), wobei das rationale Denken für eine unerlässliche Bedingung einer rationalen Handlung gehalten wird. Einen Ausdruck von Rationalität sieht man in der Haltung nur solche Sätze zu akzeptieren, die wahr sind. Die Rationalität der Welt als ihre Intelligibilität bedeutet unter anderem, dass sie sich rational erkennen lässt. Das Prinzip der Rationalität (Intelligibilität) des Seins (der Welt) hat immer viele Gegner gehabt. Ein wichtiges Problem bleibt die Universalität des Begriffs der Rationalität: Dabei wird bezweifelt, ob es eine allgemeine Rationalität gibt, die alle Fälle umfasst, in denen von einem rationalen Denken und einer rationalen Handlung die Rede ist. Ausserdem ist es problematisch, ob die Definition der epistemischen Rationalität zugleich effektive Kriterien einer Unterscheidung der rationalen von irrationalen Meinungen und Handlungen liefert. Ein Minimum an Rationalität schließt die Möglichkeit aus, zugleich p und nicht-p zu behaupten.

Rationalität als kritische Haltung (K. R. Popper) bedeutet, dass sich sowohl einer, der eine Aussage ohne genügende Beweise (Begründung) akzeptiert, irrational verhält, als auch einer, der sie ohne ausreichende Gegenargumente negiert. In der pragmatischen Zweck-Mittel-Rationalität, einer Qualifikation, die den Werten und (sittlichen) Wertsystemen eines Subjekts zukommt, geht es um eine geschickte Wahl der zu einem Zweck geeigneten Mittel (M. Weber). Eine Handlung wird für rational gehalten, wenn sie mit Rücksicht auf ein Ziel erfolgreich ist. Eine Handlung kann objektiv erfolgreich sein oder subjektiv im Bewusstsein des Handelnden (Szaniawski 1994, 535). Was die Rationalität

der Handlung betrifft haben wir heute manche Illusion der Aufklärung aufgegeben: wir wissen, dass man rational auch unwürdige Ziele verfolgen kann, z.B. auf rationale Weise Menschen töten kann. Gelegentlich wird das Kriterium der pragmatischen Rationalität dem Kriterium der epistemischen Rationalität übergeordnet: eine Handlung, die logisch irrational erscheint, kann trotzdem im pragmatischen Sinne rational sein. Wie die Diskussionen in der gegenwärtigen (analytischen) Erkenntnistheorie zeigen, ist die Identifikation der Rationalität einer Meinung (*belief*) oder einer Handlung (*act*) mit ihrer Rechtfertigung nicht problemlos. Generell geht es darum, dass das Akzeptieren von Meinungen, die nicht durch ausreichende epistemische Gründe unterstützt werden, eines vernünftigen Wesens unwürdig ist. Im Einzelnen geht es um eine befriedigende Explikation der internalistischen oder externalistischen Rechtfertigungsbedingungen. Zwar ist rational eine wahre Aussage zu akzeptieren und eine falsche abzulehnen, aber eine rational akzeptierte Aussage kann immerhin falsch und eine irrational akzeptierte Aussage kann wahr sein.

D. Rationalität von Religion.

1. Der Begriff der Rationalität von Religion. Die Anwendung des analogen Begriffs der Rationalität auf ein so komplexes Gefüge wie Religion wirft Fragen auf, die keine direkte Antwort erlauben. Was für eine Eigenschaft – wenn überhaupt – ist die Rationalität von Religion und worauf bezieht sie sich genau oder soll sie sich beziehen? Hat die Religion vielleicht eine rationale Struktur aber einer eigentümlichen Art? Im Folgenden neige ich zu der These, dass direkt nicht Religion, sondern ein Mensch rational sein kann, indem er religiös denkt und handelt. Ich unterscheide demnach zwei Hauptanwendungen des Begriffs Rationalität bezogen auf Religion: zum einen bezogen auf die (religiösen) Meinungen (epistemische Rationalität) und zum anderen auf die (religiösen) Handlungen (pragmati- sche Rationalität). Somit ist die Rationalität von Religion eine Art Verzahnung von theoretischer und praktischer Rationalität.

Der Begriff der Rationalität von Religion, der als ein verkürztes Denken funktioniert, lässt sich nicht nur eindeutig definieren, sondern er erlaubt zugleich mehrere Interpretationen. Ist die Religion rational fundiert und wenn ja, worauf: auf der Autorität des sich offenbarendes Gottes? Was ist rational? Religion im Allgemeinen oder eine konkrete Religion? Die Religion im Ganzen oder ihre doktrinäre oder ihre moralische Seite? Religion als eine Kultur- und Lebensform, eine religiöse Praxis oder eine Institution? Ein Mensch, der religiös denkt und handelt? Wie kann ein globales Rationalitätskriterium von Religion erstellt werden, wenn eine (christliche) Religion eine absolute Anerkennung fordert und ihre Wahrheitsansprüche sich *in toto* auf die ganze Wirklichkeit beziehen? Ein besonderes Problem bei der Bestimmung des Begriffs der Rationalität von Religion ist die Rationalität der religiösen Autoritäten: Gott, Propheten, Mystiker, Priester, heilige Schriften (z.B. die Bibel), Kirchen und ihre Theologen.

Zunächst ist Rationalität keine übliche Qualifikation von Religion. Sie lässt sich nicht pauschal einer Religion, jeder Religion und jeder Dimension einer Religion in gleichem Maße zuschreiben. Notwendigerweise sieht die Rationalität von Religion anders aus im Falle einer monotheistischen oder einer polyteistischen Religion, der Weltreligionen, der Offenbarungsreligionen, der Schriftreligionen und der tradionellen (*tribal religions*) Re-

ligionen der Naturvölker wie Animismus oder Ahnenkult etc., um nur einige Typen von Religion zu nennen. Unter Rationalität von Religion wurde u.a. verstanden: die historische Wahrheit einer Religion, die Begründung der Gottesexistenz, die Glaubwürdigkeit von Lehrsätzen, Authentizität, Absolutheit und existenzielle Bedeutsamkeit von Religion, die objektive Signifikanz, Widerspruchslosigkeit und Begründbarkeit religiöser Sätze (Glaubenswahrheiten), die Rechtfertigung religiöser Meinungen und religiöser Handlungen. Wie Religion gerechtfertigt werden kann, hängt davon ab, was gemeint ist: eine religiöse Aussage (*fides quae creditur*), eine religiöse Meinung (*religious belief*) oder eine religiöse Haltung (*fides qua creditur*). Sieht man die Rationalität von Religion in der Möglichkeit ihrer epistemischen Rechtfertigung, dann sind natürlich nicht alle Religionen gleichermaßen rational.

Wird Religion (klassisch) als eine Relation zwischen dem Menschen und dem *Sacrum* (einem persönlichen Gott) definiert, zerlegt sich das Problem der Rationalität von Religion u.a. in zwei Fragen: verhält sich Gott rational, wenn Er eine Verbindung mit dem Menschen eingeht? Rationalität als ein Attribut Gottes (höchste Intelligenz) anerkennen z.B. das Christentum, der Islam und das Judentum. Benimmt sich ein Mensch rational, wenn er mit Gott Kontakt aufzunehmen versucht? Beide Fragen präjudizieren gewissermaßen ihre Antwort: gibt es einen Gott (Problem der Existenz Gottes), der den Menschen geschaffen und ihm das ewige Heil versprochen hat, dann ist es für den Menschen rational, diesen Gott in irgendeiner Form von Religion zu respektieren und zu ehren.

Unter historischer Wahrheit einer Religion wird u.a. ihre Genese, die Geschichtlichkeit ihres Stifters und seiner Lehre, die Authentizität und die Glaubwürdigkeit der Heiligen Schriften, die doktrinäre und institutionelle Identität durch die Geschichte und die Kontinuität der religiösen Tradition verstanden. Besonders das Christentum ist, wie keine andere Religion, auf die geschichtliche Wahrheit ihrer historischen Ereignisse (wie die Geburt und Auferstehung Christi) angewiesen. Auf der anderen Seite ist jedes Wunder in den Evangelien in diesem Sinne irrational, weil es im Konflikt mit den Gesetzen der Natur und des logischen Denkens steht und sich rein wissenschaftlich nicht erklären lässt. Aber alle großen Religionen berufen sich auf eine göttliche Offenbarung oder die Wunder, die sich nicht rein rational begründen und erklären lassen.

Eine traditionelle (aufklärerische) Art, auf die Frage nach Rationalität von Religion zu antworten, besteht in Argumenten für oder gegen den Theismus. Die Rationalität von Religion wird also auf die Frage der Gottesexistenz und der Gottesbeweise reduziert: ist es rational an einen Gott zu glauben? Gibt es einen Gott, dann ist es auch rational, die theistische These (*theistic belief*) und die Religion zu akzeptieren. In der Frage um die Gottesexistenz sind folgende religionsphilosophische Hauptpositionen möglich: Theismus, Atheismus, Deismus, Panentheismus, religiöser Agnostizismus, Fideismus, die alle wichtige Vertreter haben. Der Theist akzeptiert die These »Gott existiert«, der Atheist akzeptiert ihre Negation und der Agnostiker suspendiert sie.

Obwohl die Standards für die Gottesbeweise (ontologische, kosmologische, teleologische) herkömmlich immer hoch gesetzt waren (A. Plantinga), haben sie verschiedene Mängel, die zeigen, dass hier die theoretisch-rationalen Argumente allein nicht ausreichen (F. v. Kutschera). Die Gottesbeweise geben nie eine absolute Gewissheit, dass Gott wirklich existiert und nötigen den Verstand nicht zwingend zur Zustimmung zu dieser These. Es wäre naiv zu erwarten, dass sich der Satz »Gott existiert« isoliert überprüfen

ließe. Er ist ein Teil eines Systems von Aussagen, das sich nur als Ganzes testen lässt. Übrigens, ist nicht in erster Linie das praktische Ziel der Gottesbeweise Ungläubige von der Gottesexistenz zu überzeugen, sondern dem Glauben, der zuerst meist Autoritäts- oder Erlebnisglaube ist, eine begrifflich nachprüfbare Grundlage und Rechtfertigung zu geben (M. Rast 1976, 154).

Der Schluß von der Rationalität der theistischen These auf die Rationalität von Religion ist nicht voll gerechtfertigt, weil sich die Rationalität von Religion nicht auf die Rationalität von Gottesbeweisen reduzieren lässt. Die meisten Gläubigen sind religiös, ohne zwingende (philosophische) Beweise für die Gottesexistenz zu haben. Für sie ist Religion immer viel mehr als eine bloße (theoretische) Anerkennung der Existenz Gottes. Nebenbei bemerkt: die Frage nach der Rationalität von Religion wird hier primär auf die monotheistischen Religionen bezogen.

2. Rationalität als Begründung religiöser Aussagen. Es gibt den Standpunkt, der die Rationalität von Religion mit ihrer Wahrheit gleichsetzt. Das heißt u.a., dass eine Religion rational ist, wenn ihre religiösen Sätze (Glaubenswahrheiten) sinnvoll und wahr sind in diesem Sinne, dass sie sich auf eine Wirklichkeit beziehen, von der behauptet wird, dass sie existiert und so oder so beschaffen ist (*objektive* Wahrheit der Religion), überdies, dass diese Sätze widerspruchsfrei sind und ein kohärentes System von Aussagen bilden.

Ob es richtig ist, die Rationalität von Religion auf die Wahrheit ihrer Glaubenssätze zu reduzieren und die von den Epistemologen diskutierten allgemeinen Kriterien der Rationalität ohne weiteres auf die religiösen Aussagen anzuwenden, hängt u.a. davon ab, ob sich in dem (heute wissenschaftlichen) Ideal des Wissens ein Platz für religiöses Wissen findet und ob es wirklich „gute Gründe" gibt, die religiösen Sätze als wahr zu akzeptieren. Objektiv rational sind für jemanden religiöse Aussagen, wenn er sie begründen, d.h. Gründe für die Glaubenswahrheiten in einer objektiven Welt finden kann. Was und wann als dieser Grund gilt, wird von den Religionsepistemologen erforscht. Spezifisch an den religiösen Lehrsätzen ist eben, dass sie eine volle auf den Glauben gestützte Gewissheit verlangen, während der Grad ihrer epistemischen Begründung niedrig bleibt, und nach den Kriterien der epistemischen Rationalität keine Glaubenswahrheiten – weder direkt noch indirekt – zureichend in einer objektiven Wirklichkeit begründet sind.

Unter den Religionsphilosophen (J. Hick) findet man die Ansicht, dass die klassische Definition der Wahrheit keine Anwendung auf die Frage nach der Wahrheit von Religion findet, weil die Sphäre des Faktischen, auf die sich die religiösen Sätze beziehen, empirisch nicht erfassbar sei. Außerdem bleibt die Objektivität, am Beispiel der Wissenschaft als Intersubjektivität (als das, was von den Wissenschaftlern akzeptiert wird) verstanden, ein vager Begriff, für den es keine eindeutigen, sachlich begründete Kriterien seiner Anwendung gibt. Bezogen auf die Religion hat er zu Folge, dass man auch in Sachen der Glaubenswahrheiten den Ansichten der Mehrheit folgen soll. Es gibt auch nichtkognitive Deutungen der religiösen Sätze, nach denen sie ihrer Natur nach einer rationalen Begründung unzugänglich sind. Hierhin gehören sowohl diese Denker, die die Offenbarung für eine unabhängige Erkenntnisquelle halten, wie auch die logischen Empiristen, für welche die religiösen Sätze keinen kognitiven Gehalt haben, sondern Ausdruck von emotional-volitiven Haltungen und Gefühlen sind, moralische Sätze oder Appelle an den Hörer sind, und daher keine sinnvollen Behauptungen, die wahr oder falsch, beweisbar oder wi-

derlegbar sein können.

Demgegenüber stellt der religiöse Rationalismus ein minimales Kriterium der Rationalität von Religion als Annahme des Gesetzes des Widerspruchs auf: die Religion ist in diesem Sinne rational, dass sich die religiösen Aussagen nicht im Widerspruch zu den Regeln des logischen Denkens befinden und ein widerspruchsfreies, kohärentes System bilden. Darauf ist jedoch zu sagen: erstens keiner Religion ist es gelungen, ihre Lehre als ein kohärentes, widerspruchsfreies System von Glaubenssätzen darzustellen; zweitens gibt es keine Möglichkeit, alle logischen Konsequenzen einer Religion zu überblicken, um eine Kontradiktion auszuschließen; drittens ist keine Religion in jeder Hinsicht mit dem wissenschaftlichen Weltbild ihrer Zeit vereinbar.

3. Rationalität als Rechtfertigung der religiösen Meinungen. Gemäß dem epistemischen Begriff des Wissens als wahre und gerechtfertigte Meinungen (*justified true beliefs*) geht es in der Frage um die Rationalität von Religion um die Bedingungen, unter welchen es für einen Menschen rational (*subjektive* Wahrheit der Religion) ist, die Glaubenssätze einer Religion in gerechtfertigter Weise zu akzeptieren. Entsprechend dem Postulat des Kritizismus (J. Locke) haben vernünftige Wesen die Pflicht, alle ihre Meinungen ohne Ausnahme einer rationalen Prüfung zu unterwerfen. Wie der oft zitierte Spruch von W. K. Clifford (1879) sagt, ist es immer, überall und für jeden falsch etwas ohne genügende Beweise zu glauben (*believe*). Fallen alle Meinungen unter eine rationale Kontrolle, gibt es keinen Grund für den Bereich der religiösen Meinungen eine Ausnahme zu machen (Immunität zu gewähren).

Nach dem klassischen Konzept der Rationalität ist es rational, eine Meinung nur dann zu akzeptieren, wenn sie gerechtfertigt ist, d.h. wenn es Gründe – epistemische oder andere – zum Akzeptieren der religiösen Meinungen gibt (*good reasons approach* – S. Toulmin, K. Baier). Das Problem verschiebt sich also auf die Frage nach den „guten Gründen". Antiszientistisch eingestellte Religionsphilosophen (W. P. Alston) verteidigen die Rationalität der religiösen Meinungen mit der Behauptung, dass die Religion selbst ihre eigenen Kriterien der Rationalität bestimme. Der Gläubige hat für seine religiösen Aussagen (und Handlungen) Gründe, die einem Ungläubigen unzugänglich bleiben.

4. Rationalität als Rechtfertigung der religiösen Handlungen. Für einen religiösen Menschen ist seine Religion immer viel mehr als eine Theorie oder ein System von religiösen Sätzen oder Meinungen, die sich objektiv rechtfertigen lassen, sondern vor allem eine besondere Lebensform, die sich in religiösen Handlungen verwirklicht. In welchem Sinne lassen diese sich als rational bezeichnen? Handelt rational der, der sich für die Religion generell (Theismus und Deismus kontra Agnostizismus oder Atheismus) oder für eine konkrete Religion entscheidet? Was für eine Rationalität kann man z.B. Christen zuschreiben, die an die Auferweckung der Toten glauben und dass die Worte und Gesten eines Priesters Sünden lossprechen und die Präsenz Gottes während der Konsekration herbeiführen können? Verhalten sich Christen rational, die sich auf den Philippinen am Karfreitag ans Kreuz schlagen lassen?

Einige Philosophen (M. Weber, logische Empiristen, kritische Rationalisten) ziehen in Zweifel, dass sich praktische Orientierungen (Zwecksetzungen und Handlungsregeln) im rationalen Sinne rechtfertigen lassen. Argumente für oder gegen praktische Orientierungen reduzieren sich letzten Endes auf subjektive Entscheidungen und Werthaltungen. Aus einer subjektiven Entscheidung für die Religion folgt natürlich nicht, dass ihre Glaubensinhalte auch objektiv wahr sind. Rational – objektiv oder subjektiv je nach der Art der Rechtfertigung – ist für jemanden eine (religiöse) Handlung, wenn er für sie gute Gründe (Motive) hat. Was es heisst, „gute Gründe" für eine Handlung zu haben, wird in der logischen Entscheidungstheorie und Wahrscheinlichkeitstheorie festgesetzt, wo andere Rationalitätskriterien gelten als bei der Rechtfertigung der epistemischen Meinungen. Dabei soll der Unterschied zwischen epistemischen und pragmatischen Gründen nicht vergessen werden. Eine religiöse Handlung, die aus epistemischen (objektiven) Gründen nicht rational ist, kann es dennoch aus pragmatischen (subjektiven) Gründen sein. Eine pragmatische Rechtfertigung der Religion als eigener Lebensform im Bereich des menschlichen Lebens bedeutet, dass sie Weltbilder vermittelt, wichtige Antworten auf Lebensfragen und somit Orientierungshilfe im Leben gibt. Bewährung im Leben ist zwar ein vages, kann aber ein vernünftiges Kriterium der Rationalität von Religion sein (F. v. Kutschera 243).

5. Wahl der Religion. Stellen wir die Frage: geschieht die Wahl einer Religion deshalb, weil sie rational ist oder spielen da andere Gründe eine wichtigere Rolle? Wer und wann verhält sich rational: ein Atheist, der die Religion ablehnt oder ein Theist, der sie bejaht? Was ist mehr rational: ein glaubender oder nicht glaubender Mensch zu sein? Sich für die Religion der Vernunft (*religio rationalis*) oder für eine der Offenbarungsreligionen zu entscheiden? Gibt es eine Möglichkeit der rationalen Wahl zwischen den großen Weltreligionen und hunderten kleineren Religionen, die trotz gravierender Unterschiede in Theorie und Praxis behaupten, dass sie ein und denselben Gott verehren? Andererseits, kann man das Verhalten eines Menschen als rational bezeichnen, wenn er die so wichtige Entscheidung, wie die Wahl einer pro- oder antireligiösen Haltung, auf die Zeit verschiebt, wenn er alle notwendigen Daten und Argumente in der Hand hat? Erwähnenswert ist in diesem Zusammenhang, dass eine allgemein pro-religiöse Haltung noch nicht gleichbedeutend mit der rationalen Wahl einer konkreten Religion oder Form der Religiosität ist.

Die Ansicht, es sei rational eine Religion zu wählen, die wahr ist, dominiert zwar. Doch vom Standpunkt eines Atheisten, der die religionsphilosophische Position eines religiösen Relativismus bezieht und die Rationalität von Religion mit der (empirischen, positiven) Wahrheit ihrer Lehre identifiziert, erscheinen alle Religionen gleich irrational: weil sie alle falsch seien, gäbe es keine Möglichkeit einer rationalen Entscheidung zwischen ihnen. Wie die Biographien vieler Atheisten zeigen, ist die anfängliche Religiosität kein zwingender Grund, ein Leben lang religiös zu bleiben. Andererseits bringen die Meinungsumfragen solche paradoxe Situationen ans Licht, dass Leute, die einer Religion angehören, nicht selten auf den ersten Blick eine irrationale, willkürliche Auswahl dieser oder jener Glaubenswahrheiten treffen, so dass jemand, der sich als Christ bezeichnet, gleichzeitig das Leben nach dem Tode und andere wichtige christliche Dogmen abstreiten kann.

Motive, warum ein Mensch religiös ist oder nicht, sich für diese oder jene Form der Religiosität entscheidet, sind vielfältig und verschiedenartig: bewusste und unbewusste, objektive und subjektive, rationale und nichtrationale, kulturelle, soziologische und psy-

chologische, so wie viele andere Motive menschlicher Entscheidungen. Somit fällt die Wahl einer Religion in den Bereich der Entscheidungs- und Handlungstheorie, die formale Modelle der menschlichen Handlung konstruieren. Eine Religion wird selten aus rein rationalen Gründen durch einen vereinzelten, bewußten Akt der Vernunft gewählt und die Mehrheit der Glaubenden hat es nie systematisch versucht irgendwelche rationale Argumente für ihre religiöse Haltung zu finden. In den meisten Fällen werden sie in eine Religion – als Lebensform – „hineingeworfen" durch die Geburt in eine religiöse Kultur und Tradition, sowie die Erziehung in einer religiösen Familie. Die Tatsache, dass man Anhänger religiöser Meinungen und Praktiken findet (der Kasus der Magie von Azande), die gemäß der wissenschaftlichen Vernunft zutiefst irrational erscheinen, beweist, dass hier andere als rationale Faktoren mitspielen.

Wo eine Religion bewusst gewählt wird, spielen meistens außerepistemische, pragmatische Gründe die Hauptrolle, primär die soteriologische (heilbringende) Funktion der Religion, obwohl es wünschenswert wäre, dass die epistemischen und pragmatischen Gründe im Einklang stünden. Es kommen solche Qualifikationen der Religion in Frage wie heilig, heilsam, glaubwürdig oder menschenfreundlich. Die Tatsache, dass die meisten Menschen nicht aus theoretischen (philosophischen) Gründen religiös sind – z.B. weil sie für sich glaubwürdige Beweise für die Existenz Gottes gefunden haben – bedeutet nicht, dass sie sich somit zwangsläufig irrational verhalten. Die Religion verspricht, obwohl unterschiedlich in verschiedenen Religionen, – irgendeine Form des Heils (der Heilszusage): ein erfülltes und glückliches Leben nach dem Tode, die Aufhebung der Grenzen irdischer Existenz, die Vollendung des Lebens in einer ewigen Seligkeit. Sie wird gesucht, weil sie wesentliche, geistige Bedürfnisse des Menschen befriedigt, die Welt – wie Philosophie und Wissenschaft – existenziell verständlich (einsichtig) macht, Antwort auf die existenziellen Fragen nach dem Sinn des Lebens (»Wer bin ich?«, »Woher komme ich?«, »Wohin gehe ich«?) und somit Orientierung im Leben gibt und die Erlösung vom Übel, von Sünde, Schuld, Leiden, Tod und Angst in Aussicht stellt. Eine andere Frage ist es, ob sie ihre Versprechen auch wirklich erfüllen kann. Weil sich diese soteriologische Dimension der Religion auf das Leben nach dem Tode bezieht, entgeht sie einer empirischen (und in diesem Sinne rationalen) Verifikation oder Falsifikation.

6. Zweifel am Begriff der Rationalität von Religion. Nicht nur unter den Atheisten, sondern auch den Theisten (religiöse Akognitivisten) gibt es Zweifel, ob die Rationalität eine geeignete Eigenschaft und ein adäquates Kriterium der Wahl einer Religion ist. Im Rahmen des Theismus findet man generell zwei, sich nicht notwendig ausschließende Einstellungen der Religion und dem Göttlichen gegenüber: die eine hält Gott – *pura intelligentia* – für reine Rationalität, die andere sieht das Numinose als das Geheimnisvolle, das radikal jenseits der Vernunft liegt.

Zu den Strategien der Verteidigung der Rationalität von Religion gehört die Metapher der Tiefe: die Religion erscheint nur in ihrer „flachen" Dimension irrational, aber in ihrer „tiefen" ist sie rational. Oder umgekehrt: die „flache" Dimension der Religion ist rational, aber was „tiefer" liegt, entgeht jedem rationalen Verstehen. Für Religion werden hier also zwei unterschiedliche Verständnisdimensionen angenommen: eine gleichsam „oberflächliche" und eine „tiefe". Je nach Argumentationsstandpunkt erscheint die „oberflächliche" Dimension zunächst durch irrationale Zuwendung als kein Verstandesobjekt zu

gelten; dringt man jedoch in ihre Tiefdimension vor, ergeben sich rational nachvollziehbare Relationen. Oder, wie der zweite Standpunkt meint, betrifft die Rationalität nur die Oberflächendimension, und alles, was „tiefer" liegt, entzieht sich jedem Verstandeszugang.

So wird behauptet, dass die religiösen Geheimnisse (Mysteria) die Vernunft zwar überschreiten, sie aber nicht unbedingt mit ihr unvereinbar (widersinnig) bleiben müssen. Die theistische, antirationalistische (fideistische) »Flucht in den Glauben«, die den Glauben für ein Wagnis und eine ausschließliche Quelle der religiösen Erfahrung hält, teilt mit den Atheisten die Ansicht, dass die mystische (eben „tiefe") Dimension der Religion prinzipiell „irrational" bleibt. Eine Konsequenz des Fideismus ist die These, dass in Bezug auf Rationalität alle Religionen gleich sind. Es wird also behauptet, dass die Rationalität kein immanenter Wert der Religion sein kann, weil das, was für eine Religion wesentlich ist – die persönliche Begegnung mit der göttlichen Macht (in monotheistischen Religionen mit einem persönlichen Gott) – nicht von rationalen Argumenten abhängig bleibt, sondern ein Akt des Glaubens ist – im Christentum eine Gottesgnade – und somit einer rationalen Qualifikation unzugänglich bleiben muss. Weil sich die Religion auf eine transzendente Wirklichkeit bezieht, hat sie keine ausreichende Gründe ihrer Existenz in einer rationalen Ordnung und der Glaubensakt *qua* Glaubensakt muss sich *ex definitione* außerhalb des Bereichs des logischen Denkens abspielen.

Ein Beispiel für die Schwierigkeiten die Glaubenssätze mit der Vernunft und der epistemischen Rationalität in Einklang zu bringen, sind die Dogmen des Katholizismus wie die Hl. Dreifaltigkeit, die jungfräuliche Geburt, die Auferstehung Christi vom Tode, oder die Präsenz Christi in der Eucharistie. Wie die christliche Theologie betont, bedeutet jedoch die Tatsache, dass die Glaubenswahrheiten die menschliche Vernunft überschreiten, nicht notwendig, dass sie in diesem Sinne irrational sind, dass sie gegen die Vernunft verstoßen.

E. Abschließende Bemerkungen.

Ich bin mir darüber klar, dass ich auf die Titelfrage, *Sind alle Religionen gleichermaßen rational?*, keine eindeutige Antwort gegeben habe, aber das war auch nicht mein Ziel. Mit meinen Ausführungen wollte ich lediglich zeigen, dass

1. erstens auf die allgemein gestellte Frage, worin die Rationalität von Religion besteht, mehrere, konkurrierende Antworten möglich sind;
2. zweitens je nach dem Begriff der Religion und der Rationalität der Begriff der Rationalität (Irrationalität) der Religion eine andere Bedeutung hat;
3. es drittens keine eindeutigen Kriterien der Rationalität von Religion im Allgemeinen oder einer konkreten Religion im Besonderen gibt;
4. viertens, der direkte Träger (als erstes Analogat) der Rationalität ist der religiöse Mensch, der Religion dagegen kommt diese Funktion nur indirekt zu; das, was als Rationalität von Religion bezeichnet wird, bezieht sich primär auf den religiösen Menschen: auf die Rationalität seiner religiösen Meinungen und auf sein religiöses Verhalten;
5. sich fünftens allein mit logischen (semiotischen) Mitteln und rationalen Gründen die Rationalität einer religiösen Meinung oder Handlung nicht rechtfertigen lässt. Es gibt an-

dere Faktoren als kognitive, von denen die Antwort auf die Frage nach der Rationalität von Religion, u.a. ob es rational ist, zu glauben oder nicht, ein Christ zu sein oder nicht, abhängig bleibt;
6 man sechstens – obwohl die Rationalität nicht die wichtigste Qualifikation der Religion ist – mit Recht fragen kann, ob die Wahl einer Religion im Allgemeinen oder einer konkreten Religion im Besonderen rational ist;
7. siebtens die Tatsache, dass man mit Berufung auf die Vernunft die Religion sowohl unterstützen wie auch kritisieren kann, praktisch bedeuten könnte, dass man sowohl aus rationalen, wie auch aus irrationalen Gründen religiös sein kann.

Literatur

Alston, William P. 1999: *The Distinctiveness of the Epistemology of Religious Belief*, in G. Brüntrup, R. Tacelli (eds.), *The Rationality of Theism*, Dordrecht: Kluwer, 237–254.

Bocheński, I. M. 1968: *Logik der Religion*, Paderborn: Bachem.

Bronk, A. 1996 *Nauka wobec religii: teoretyczne podstawy nauk o religii* (*Wissenschaft gegenüber der Religion: theoretische Grundlagen der Religionswissenschaften*), Lublin: TN KUL.

Clifford, W. K. 1879: *The Ethics of Belief*, in *Lectures and Essays*, London: Macmillan.

Hick, John (ed.) 1966: *Faith and the Philosophers*, London, New York: Macmillan.

Kutschera, F. von 1990: *Vernunft und Glaube*, Berlin, New York: Gruyter.

Lenzen, W. 1980 *Glauben, Wissen und Wahrscheinlichkeit. Systeme der epistemischen Logik*, Wien, New York: Springer.

Lukes, S. 1970: *Some Problems about Rationality*, in B. R. Wilson (ed.) *Rationality*, New York: Harper and Row, 194–213.

Plantinga, A. 1985: *Reason and Belief in God*, in A. Plantinga, N. Wolterstorff (eds.) *Faith and Rationality. Reason and Belief in God*, Notre Dame: University of Notre Dame Press, 16–93.

Plantinga, A. 1993: *Religious Belief, Epistemology of*, in J. Dancy, E. Sosa (eds.) *A Companion to Epistemology*, Oxford, Cambridge, Mass.: Blackwell, 436–441.

Przelecki, M. 1996: *Poza granicami nauki (Außerhalb der Grenzen der Wissenschaft)*, Warszawa: Polskie Towarzystwo Semiotyczne.

Rast, M. 1976: in W. Brugger (ed.), Philosophisches Wörterbuch, Freiburg, Basel, Wien: Herder, 154–155.

Religion, 1992: in J. Ritter, K. Gründer (Hg.) *Historisches Wörterbuch der Philosophie*, Basel, Stuttgart, vol. 8, 632–713.

Stenmark, M. 1995: *Rationality in Science, Religion and Everyday Life. A Critical Evaluation of Four Models of Rationality*, Notre Dame, University of Notre Dame Press.

Swinburne, R. 1979: *The Existence of God*, Oxford: Clarendon Press.

Swinburne, R. 1993: *The Coherence of Theism*, Oxford: Oxford University Press 1977.

Szaniawski, K. 1994: *Racjonalność jako wartość* (*Rationalität als ein Wert*) in derselben, *O nauce, rozumowaniu i wartościach. Pisma wybrane*, Warszawa: PWN, 531–539.

Weischedel, W. 1975: *Der Gott der Philosophen. Grundlegung einer philosophischen Theologie im Zeitalter des Nihilismus*, München: Nymphenburger Verlagshandlung.

Wittgenstein, L. 1971: *Vorlesungen und Gespräche über Ästhetik, Psychologie und Religion*, Göttingen: Vandenhoeck & Ruprecht.

Wolterstorff, N. 1999: *Epistemology of Religion* in J. Greco, E. Sosa (eds.) *Epistemology*, Oxford: Blackwell, 303–324.

Wolterstorff, N. 1983: *Can Belief in God be Rational If It Has No Foundations?* in A. Plantinga, N. Wolterstorff (eds.) *Faith and Rationality: Reason and Belief in God*, Notre Dame: University of Notre Dame Press, 135–186.

Posthumanism: Engineering in the Place of Ethics

STEPHEN R.L. CLARK

1. Folk Remedies and Modernity

My purpose is to examine the prospects for an 'engineering' or 'biotechnological' solution to those failures of social and personal adaptation which we ordinarily call 'wickedness' or 'vice' (see also Clark 1995, 1999, 2000a, 2000b). Ethics, like medicine, used to provide folk remedies and placebos for human ills, and especially the hope that moral discipline could lead us, in the end, to Eden or the Celestial City. Some have hoped that early conditioning by loving parents would really be sufficient; others that 'education', by professionals, could show us how to achieve the goals, the goods, that all of us, of course, must 'really' want. One long neglected tradition of ethical thought has suggested that even our virtues are but 'noble vices', and that what is needed is a radical reconstruction of human and animal nature which we cannot ourselves encompass, but philosophical optimists have usually preferred to believe that some combination of loving care and intellectual rigour will produce the kind of people that are needed for the common good. Less philosophical optimists have relied instead on aphorisms, scornful looks and more or less high-minded bribes for good behaviour.

Perhaps those moral disciplines were the best that we could manage, and perhaps some will prove to have had more sense in them than we now easily suppose. If there is a case for leeches, then perhaps there is also a case for sages, and even for repressive teachers! But it is as well to recall the Buddhist story: upon meeting a sage whose ascetic discipline over twenty years had enabled him to levitate across a river, the Buddha mildly retorted that it was cheaper to take a ferry. The mathematical and lexicographical calculations that once took scholars decades to perform, and required an intellectual and moral discipline that had its own personal costs, can now be handled by a chip. Moralizing advice to 'pull yourself together' is almost always likely to be less successful than a pill. Once upon a time we had to put up with rain, and loving parents encouraged their children to endure it gladly. But then we invented houses, raincoats and umbrellas. Once upon a time (and now) we had to put up with *pain*, and loving parents taught us ways of coping. But then we invented analgesics. Once upon a time, the perils of sexual activity were so great that it required regulation. But then we discovered antibiotics and (fairly) effective contraceptives. Once upon a time (and now) heroin, nicotine or alcohol addiction could be handled only by 'moral effort', public confession and clumsy attempts to provide slightly less harmful alternatives. Once upon a time we needed competent and intrepid cavalry, and practised the arts of war by hunting foxes. Once upon a time the principal way of dealing with dangerous or obnoxious people was to expose them to public scorn, in the hope that they would internalize a secret censor. Once upon a time we desperately needed *moral* courage, and 'Dutch courage' (that is, gin) had too many side effects. New understanding of our limbic (and other) systems, and our concomitant ability

B. Brogaard, B. Smith (Hg.), *Rationality and Irrationality / Rationalität und Irrationalität*, S. 62–76.

to construct *new* systems, make all the old ways obsolete. Now engineering will do more than ethics to achieve our goals.

Biochemical or prosthetic aids are patches on the problems. Sometimes they may merely displace the problems: chemical castration may prevent one sort of violence, but prepare the way for other cruelties. The biochemical tools we need will have to be targetted carefully, at the brain much more than at the glands. No doubt their use will also have to be accompanied by verbal therapies of one kind or another. The elderly Sophocles may have been glad that he was no longer burdened by sexual desire (according to Plato *Republic* 1.369): other elderly males (or females) may regret the loss, and seek ways to compensate for their own lack of lust or lustiness. But the converse is also true: even if we are genuinely convinced that we should not indulge in this or that behaviour, most of us find it very difficult to put it from our minds. Chemical assistance may actually achieve what moral discipline does not. If we want *not* to want certain things, and want to want others, then why not extinguish or create such wants by any practicable means? Some critics may suppose that it is – somehow – *better* to use moral and mental disciplines than to use chemical aids to equanimity: but why? Our present drugs, no doubt, are inefficient and have unwanted side-effects – but we all willingly use analgesics rather than self-control and careful redescription to cope with pain. Why not use anaphrodisiacs (or conversely, aphrodisiacs)?

It may also prove to be possible to *eliminate* the problems by genetic action. Genetic therapy strictly so-called – the addition of new genes to existing, developed organisms – is no more than a dream at present. Even if the new gene is taken up by the appropriate cells, the effect is more likely to be lethal. The most that most 'genetic engineering' manages, so far, is to recognize an inappropriate gene, and either abort the unfortunate foetus or, in a few cases, prepare the appropriate diet. Adding genes even to newly fertilized zygotes is a chancy affair: in experimental animals, almost all the additions fail, and we are not so far gone in moral corruption as to attempt that level of waste with *human* victims.

But our present failures are likely to lead on to success. Some day we shall be able to read and amend the genetic codes of people. Now that the first stage of the Human Genome Project is complete we can begin to identify the effects of deleting or of adding this or that – at first in mice and in macaques, and later, watchfully, in human embryos. We shall be able to replace dangerous alleles, and add alleles that typically result in phenotypic variations that we want. In all but the simplest cases, to be sure, there is no exact and one-for-one relationship between genotype and phenotype – but the *potential* may be enough for us to engineer, and we may, as our knowledge grows, discover how to guide a genotype to its desired destination. Some critics seem to think that only an organism's gross anatomical features can be affected by genetic change: I can see no difference in principle between what *cells* are programmed to do, and what whole organisms are programmed to do. Behavioural patterns are as inheritable as anatomical. Nor can I see any reason in principle why *human* organisms should be unlike all others: there is ample reason to believe that we inherit our capacities, even if the particular shape that those inherited patterns take is determined by local context and culture. We walk and talk and interact with others against all odds, and in remarkably similar ways. It would be very odd indeed if *all* that we inherited were patterns that we shared with every other human being: variation is as much a biological reality as identity. And insofar as any individual human

being is as much a product of her own action as of her inheritance, that fact also has genetic roots. The fact that we *can* 'make ourselves' (whereas other creatures, probably, are far more constrained) is something that, in principle, we could edit out of selected lines, or else constrain in supposedly useful ways.

It is far too easy a riposte to claim that any attempt to emphasise genetic factors requires 'genetic determinism': no one need suppose that genes operate in a vacuum, nor even that their effects are always easy to predict. Obviously, the genes with which we shall be concerned require specific cellular, uterine and extra-uterine environments for their very various effects to be made manifest. Obviously all developed organisms have themselves a part to play in their own self-construction. Equally obviously, they can do none of these things without the appropriate genetic heritage. So far, we have only noticed *simple* cases (where a particular allele, in an ordinary environment, has a well-defined phenotypic effect). So far, we have only concerned ourselves with *obviously* harmful phenotypic features (cystic fibrosis, or Tay-Sachs Syndrome). But our knowledge will increase, and the range of undesirable and desirable phenotypes to be considered. Alcoholism, autism and Asperger's Syndrome may have identifiable genetic features, and be thought worth counteracting. So may nerviness, or xenophobia, or adrenalin-addiction.

In C.J.Cherryh's *Cyteen*, the 'architect of Union', Ari Emory, speaks as follows:

> My language is partly mathematical, partly biochemical, partly semantics: I study biochemical systems – human beings – which react predictably on a biochemical level to stimuli passing through a system of receptors – hardware – of biochemically determined sensitivity; through a biochemical processor of biochemically determined efficiency – hardware again – dependent on a self-programming system which is also biochemical, which produces a uniquely tailored software capable of receiving information from another human being with a degree of specificity limited principally by its own hardware, its own software, and semantics. (Cherryh 1989: 221)

In Cherryh's universe the techniques for identifying genesets and psychsets, dependent on the interaction of geneset and environment, are subtle and powerful. Whether their use can be compatible with anything we ordinarily conceive as human dignity is one problem. What they could or should be used to gain is another.

2. The Circularity of the Human Predicament

For why should any particular destination be desired? E.O.Wilson remarks that the time is coming when we shall have to decide how human we wish ourselves to be (Wilson 1978: 208). Once we know how our capacities, emotions and value judgements are produced, we can decide how much of what we, mostly, are is no longer relevant.

> Human nature is a hodgepodge of special genetic adaptations to an environment largely vanished, the world of the Ice-Age hunter-gatherer. ... We are forced to choose among the elements of human nature by reference to value systems which these same elements created in an evolutionary age now long vanished. *Fortunately, this circularity of the human predicament is not so tight that it cannot be broken through an exercise of will* (Wilson 1978: 196) –

though he concedes on an earlier page that 'it would be premature to assume that modern civilizations have been built entirely on genetic capital accumulated during the long haul of the Ice Age' (Wilson 1978: 88). Whenever exactly that capital was acquired, the characters that gave our ancestors the edge against their immediate cousins may not be those we can now afford. Male adolescents, for example, have always had to form alliances outside the central or domestic scene within which breeding occurs. Only those who survive, and prosper, ever return to breed – and the traits that once were needed for success may have no obvious relevance to the present needs of individuals, or groups, or lineages. No doubt those traits were never as merely brutal as folk-ethnography supposes. Charm, beauty, wit, affection, loyalty and skill won as many allies and mates as violence (and so did the capacity to imagine *and communicate* new worlds of imagination) – except that particular human troups, or the male fringes of those troups, most probably attacked their neighbours on occasion, killing the males and raping the females. And then devised emotional strategies to reconcile their aggression and their affection, or at least tolerance, for the resultant cubs. Successful males, as they grew more Macchiavellian, sought ways to disarm and divert potential rivals. The point of war is not *only* to acquire more territory and more goods: it is to give aggressive males someone else to fight than the existing rulers. None of these features, by the way, need be particular to males, even if it was through natural selection amongst males that they were predominantly fixed in our genome, any more than those features chiefly selected amongst females need now be tied to femaleness.

Some critics – it is worth pausing to observe – appear to think that explaining the biological roots of male aggression must amount to licensing that aggression. It should be obvious that I have no such intention. But it is also obvious that some theorists have allowed the inference. Behavioural patterns exist across the species (and much of the biological family) because in Darwinian terms they paid. 'Human behavior – like the deepest capacities for emotional response which drive and guide it – is the circuitous technique by which human genetic material has been and will be kept intact. Morality has no other demonstrable function' (Wilson 1978: 167) – that is, or so Wilson declares, there is nothing else that it consistently does, and that explains the particular shape it takes.

Maybe this can be altered by a collective 'act of will'. But to what end? If everything *can* be remodelled, why and in what direction? Edmund Burke, it seems to me, was entirely right to fear the plans of self-selected constitutional lawyers. 'When antient opinions and rules of life are taken away, the loss cannot possibly be estimated. From that moment we have no compass to govern us; nor can we know distinctly to what port we steer' (Burke 1968: 172; see also 129ff). Why should genetic engineers, or politicians, or philosophers do better? And what could 'better' mean?

Why should 'we' struggle, for example, to heal sickness or deformity if it is easier (as it may be) to reconcile the sick or disabled to their condition? What is the point of feeding desires that cannot now, or ever, be met for everyone, if it is also possible to *construct* desires, in suitable populations, that *can* be met? As C.S.Lewis pointed out, the power of Man over Nature is always, in reality, the power of some men over others (Lewis 1946;

see also Lewis 1945). Once we understand more of the routes from DNA to phenotypic effects, even our biochemical controls will be more effective – and we could afford to keep our psychopaths and schizophrenics as variants within the range? These variants, after all, have probably had advantages in the past, and may still occasionally be needed, if only by our rulers.

Wilson's way of breaking into the circle he describes is simply to declare – by an act of will – that our aim *must* be to facilitate the survival of 'human genes'. 'The individual is an evanescent combination of genes drawn from [the human gene-pool], one whose hereditary material will soon be dissolved back into it.' (Wilson 1978: 197). But what justifies or compels our thinking *that* goal unalterable? When Lewis described the planetary angel Malacandra's conversation with the corrupted scientist Weston, he can hardly have imagined that anyone would so readily admit the charge:

> You do not love any one of your race ... You do not love the mind of your race, nor the body. Any kind of creature will please you if only it is begotten by your kind as they are now. It seems to me ... that what you really love is no completed creature but the very seed itself; for that is all that is left.

Weston retorts by appealing to 'a man's [fundamental] loyalty to humanity', and the angel continues:

> I see now how the lord of the silent world has bent you. There are laws that all *hnau* [sc. 'rational animal'] know, of pity and straight dealing and shame and the like, and one of these is the love of kindred. He has taught you to break all of them except this one, which is not one of the greatest laws; this one he has bent till it becomes folly and has set it up, thus bent, to be a little, blind Oyarsa [sc. 'angelic ruler'] in your brain. And now you can do nothing but obey it, though if we ask you why it is a law you can give no other reason for it than for all the other and greater laws which it drives you to disobey. (Lewis 1952: 163)

If the laws of pity, shame and straight dealing (and other laws like honesty and intellectual rigour) are only useful adaptations, then yet other laws may also, in their turn, be 'useful'. Maybe the most useful might indeed be simply to do whatever has, as it presently seems to us, the largest chance of increasing our posterity – but why should *that* be our goal any more than to do justice and love mercy? Most of our genes, for that matter, can be found elsewhere, in whatever interesting combination. Why should it matter whether they persist in populations literally descended from the present human one or not?

Some have thought that fundamental impulses are immune to deconstruction. 'Parents are no more likely to stop loving their children once they understand the role that such feelings play in the perpetuation of their genes than they are to cease enjoying orgasm once they understand its evolutionary role.' (Hull 1989: 258). But it is implicit in this claim that parental love and orgasm are just natural events, owing nothing to the beliefs and projects of particular human beings, and without any cognitive significance. All Dante *really* wanted was to get Beatrice into bed, and pregnant, and nothing would have altered in his heart if he'd come to believe that Beatrice's beauty was no more than glamour. Romantic love is easily deconstructed: why should we suppose that we will feel it still once we have realized its origin, or feel it as anything but a passing dramatic fancy, or a sickness to be alleviated? Parental love is also, in large part, an artifact. Consider it a device, deployed by infants, to secure themselves: the unfortunate birds whom cuckoo-chicks control would feel quite differently if they could wake up. Is parental love, even as a 'natural' feeling, so secure? Parents have been known to shake or strike their children, to abandon them or kill them, without obvious qualms. No doubt those patterns of behaviour too can be made to seem the 'fitter': better unload unsatisfactory kids and start again, especially if they might not be ours. Even if we manage to endure the costs of caring, there will come a moment when we wish them gone. Is it obvious that sacrificial care will long outlast the discovery that kids are cuckoos? Even before we noticed, our care was less than perfect: knowing that the passing enjoyment of parental love is only a trick to keep us careful of our genes, will we really be as caring?

Darwin's declaration that 'if men were reared under precisely the same condition as hive-bees, there can hardly be any doubt that our unmarried females would, like worker bees, think it a sacred duty to kill their brothers, and mothers would strike to kill their fertile daughters, and no one would think of interfering' (Darwin 1981: vol. 1: 73; see Thompson 1995: 143), was absurd, and Sidgwick's comment (cited by Darwin) that 'a superior bee would aspire to a milder solution of the population problem' (Sidgwick 1872: 231) at least apt. How could they *not* question their own motives, and feel less constraint in following those paths? The implicit claim that we *cannot* question our own most deeply rooted impulses, and will not even think to do so, is refuted daily. We are repelled by obvious sickness or deformity, eager to attach ourselves to eminent persons (and eventually to displace them), sexually attracted by those who are different, but not too different, from ourselves, devoted to our children until the moment when they get too much on our nerves. All these impulses can be given an evolutionary explanation, and will soon be identified with particular biochemical states. Is it *obvious* that no-one has ever challenged them, or that we could not, even now, override them? Is it *obvious* that we should override them only in the hope of having more descendants if we change? That desire too is biochemical, and might as well be deleted.

3. An Aristotelian Digression

I am well aware that Wilson's own image of philosophical ethics – that it amounts to philosophers' consulting their limbic systems for particular judgements – is naïve. Few moral philosophers are crude intuitionists, convinced that their own gut-feelings and immediate 'moral impressions' constitute a sound basis for action and advice. Standard moral theorists prefer to ask, with whatever subtle qualification, what sort of action would have the 'best' consequences for all those affected, or what would be agreed by all those affected, or what would be required by a well-informed and unbiased intelligence with an interest in the interests of all those affected. Even consequentialists, in practice, seek to construct rules to guide our action. Other theorists attend more openly to questions of character and motivation. Some believe that Duty is a transcendental norm, binding us to act in one way or another *whatever* our particular inclinations. Others suggest

that there are real norms at work in the world, guiding everything towards a proper order, and that human beings are unusual only in that we can occasionally be aware of the attraction. None of these possibilities, however, make it unreasonable to ask what it is that we are generally inclined to do, desire, admire or feel revolted by. Nor is it unreasonable to ask how we might help to realise, or organize, or correct those inclinations.

Aristotle distinguishes six sorts of ethical character: beastliness, weakness, vice, self-control, virtue and heroism. The beastly have desires outside the usual human range, and ones that seem ungovernable by any moral effort. We are now readily persuaded that the only way of dealing with such people is to kill, incarcerate *or cure* them. We could, in principle, discover what neurological and biochemical condition it is that drives them to rape or eat children. Whether this condition is one that literally *any* of us might develop, or is instead one for which there is a genetic predisposition or destiny, hardly matters. They are desires that most of us cannot bear to acknowledge or imagine, and so seem easily open to merely physiological reconstruction. Weakness of will (*akrasia*) is more familiar, and more openly acknowledged. We all know what it is like to do what we momentarily desire rather than what, all things considered, we think best. We are easily seduced or terrified into wrong action, and correspondingly willing to reinforce our own or others' self-control by emphasising that it is up to us to control those momentary desires or fears – by redirection, recitation of appropriate moral mantras, and reminders of the appropriate penalties. As Aristotle recognized, those who are drunk – with alcohol or desire – may not be doing what they have themselves decided, but should still be held accountable for what they 'non-deliberately' do. Those who know (as anyone should know) that they would behave, in their own terms, badly if they drink, have no business drinking.

Vice, on the other hand, involves deliberate action, in the light of distorted values. The vicious feel no compunction about theft, rape or murder. Their values are unfortunately ones that any of us can understand: unlike the merely beastly, the vicious, in a sense, are only doing what any of us can easily desire. Attempting to 'cure' them, until now, could only be by showing them what they are missing (honest friendship, for example) or making it clear to them that they will be made to suffer for their crimes. The former technique may possibily encourage the growth of virtue; the latter only of self-control. We are at present disinclined to believe that 'vicious people' are born with a scale of values rather unlike 'our own'. It is more likely, perhaps, that they desire and fear exactly what we do, but have discovered or been taught a different scale. Whereas most of us, self-consciously virtuous or well-behaved, prefer peace to violence, the vicious enjoy violence more. Whereas most of us desire the respect or affection of the 'civilized', the vicious gain their self-esteem from the admiration or the terror of more corrupt societies. Crudely engineering solutions (like those of Anthony Burgess's *A Clockwork Orange*, perhaps) are probably unlikely to succeed. Indeed such solutions are themselves, for most of us, the outcome of thoroughly vicious temperaments. But it remains an open question whether we might hope to reorganize priorities by biotechnological means, and even identify the genotypes most likely to lead to phenotypes addicted to excitement, violence and a parochial outlook.

Few of us aspire to more than self-control. We recognize fears and desires that we must not, dare not, allow into the public sphere. We shame or terrify ourselves into obedience to those life-plans that seem, all things considered, to allow us to have honest

friends. As I have already suggested, it is difficult to see why we should not employ biotechnological devices to assist us. Or rather, it is easy enough to see why this has not previously been a good strategy: the only such techniques available to us have had their bad effects, including addiction and the gradual destruction of our own deliberative faculties. Keeping one's spirits up by using alcohol or nicotine may be physically harmful, and may also distort our value set to the point where we may slip from self-control to weakness and to vice or beastliness. But this is not to say that there may not be better techniques, discoverable as we learn how our brains work. Virtue (that is, the uncoerced and undivided bias towards right action) may itself be put more easily in our power. Some of those who now seem virtuous, of course, may only never have been fully tested. In peaceful days, and prosperous, we may not need much courage, temperance, good sense or even any strong sense of justice: in more calamitous times, we may be surprised to find that 'honest citizens' are – really – vicious, and that the town wastrels are better equipped to cope. Ordinary virtue is usually much less than heroic or saintly. Heroism or saintliness involves the happy sacrifice of ordinary goals for some true good. As a natural phenomenon, it is no more than an extreme version of whatever capacities for courage or compassion have been naturally successful – just as beastly or vicious people are no more than extremes at the other end of possible human (primate, mammalian) behaviour. Whether such heroism or saintliness reveals a deeper truth is a topic for another occasion (see Clark 1997). In the present context it is enough that we could, in principle, identify the natural roots even of such exceptional characters, and begin to create them.

Aristotelians, of course, are unlikely to agree with Wilson that the only proper goal for human beings is to maintain human genetic material, in whatever phenotypes. Our goal should be instead to live as human, or as nearly divine, lives as we can manage. Our image of such a divine humanity, in turn, is formed through natural processes – without thereby being an image to be discarded once we understand those processes. It is possible to assess our different images without recourse to a merely instrumental reason (as though there were indeed an obvious goal – the maintenance of our genetic material – for which there were more or less successful methods). For Aristotle, as for other such moralists, the goal must rather be to serve and contemplate the divine (Aristotle, *Eudemian Ethics* 8.1249b9ff). But that too is another story.

4. The Old and the New Breeding Programmes

Wilson's actual arguments for insisting that our only real or at least our chief aim (the standard by which in future we shall evaluate all other goals and rules) must be the survival of 'human genetic material' are plainly absurd. It does not follow that they are necessarily unconvincing. Whole civilizations have been founded upon principles that other peoples, other generations, find ridiculous. So what would it be like to take the principle seriously, and reorder our lives accordingly? Might we not reckon, for example, that our lineage will last much longer if we speciate? Wilson himself concedes or claims that diversity is 'a cardinal value' (Wilson 1978: 198). A lineage or clade will usually do 'better', over evolutionary time (that is to say, last longer), if it divides into many new species, each with a distinctive niche. In the cichlids of the East African Rift lakes, for ex-

ample, 'the two sets of jaws, fine-tuned according to food habits, allow each species to occupy its own very specific ecological niche. In this manner, hundreds of species can co-exist without directly competing. If instead these cichlids had tried to exploit the same resources, most would have been driven to extinction' (Stiassny & Meyer 1999: 48). Gause's Law states that maximum competition is to be found between those species with identical needs (Wilson 1978: 175): – so if we are really concerned about having descendants, we had better speciate to avoid destructive competition. Biological and social engineers may as easily get their satisfactions by constructing or reconstructing old humanities – the separate worlds of rich and poor, or male and female, or vagabond and stay-at-home. Or else they may invent some new humanities at least as various as hominoids. Why not? As Chesterton saw, the sub-conscious popular instinct against Darwinism had its roots in the knowledge that 'when once one begins to think of man as a shifting and alterable thing, it is always easy for the strong and crafty to twist him into new shapes for all kinds of unnatural purposes. The popular instinct sees in such developments the possibility of backs bowed and hunchbacked for their burden, or limbs twisted for their task. ... The rich man may come to be breeding a tribe of dwarfs to be his jockeys, and a tribe of giants to be his hall-porters.' (Chesterton 1910: 259)

None of this is entirely unfamiliar to traditional moralists. The breeding programme that Plato proposed in his *Republic* had one interesting difference: its goal was, precisely, to manipulate and breed 'the strong and crafty' as we have tamed wolves or oxen. Guard dogs need appropriate temperaments and capabilities: so also guardians. But in *The Statesman*, a wider vision beckons. The first attempt Plato makes to define the Statesman (*The Statesman* 258b-267d) distinguishes him as one who maintains his rule, by mental power and force of personality over a host of subordinate directors with the object of breeding and nurturing herds of tame, gregarious animals: specifically those land-dwelling, walking, hornless, and bipedal animals that we call human. The art of governing people is to manage the breeding and nurturing of a particular kind of herd animal: 'the science of the collective rearing of men as distinct from the rearing of horses or other animals' (see also *The Laws* 680e).

Maybe we do this even without deliberate plan. 'Malefactors are executed, or imprisoned for long periods, so that they cannot freely transmit their bad qualities. Melancholic and insane persons are confined, or commit suicide. Violent and quarrelsome persons often come to a bloody end. Restless men who will not follow any steady occupation – and this relic of barbarism is a great check to civilization – emigrate to newly-settled countries, where they prove useful pioneers.' (Darwin 1981: vol.1: 172) Ernest A.Hooton, and others, wished the strategy to be more ruthlessly employed. 'The elimination of crime can be effected only by the extirpation of the physically, mentally, and morally unfit, or by their complete segregation in a socially aseptic environment.'(Hooton 1939: 309; see Gould 1981: 111) So did Ernst Haeckel and his Monist League: 'The "redemption from evil" [for the ill, deformed and criminal] should be accomplished by a dose of some painless and rapid poison.'(Haeckel 1904: 118f; see Gasman 1971: 95). One problem is, as before, that the genotype, the inheritable factor, is not strictly tied to one particular phenotype (and Darwin had no evidence at all that malefactors had their character by inheritance, or that they would transmit it to their children, or that they even differed in any essential way from any random sample of the population). We select partners by their overt

characteristics, whatever their genetic character. The point is not – or not only – that we can be tricked, but that we cannot tell what other hidden effects their genes, whatever they are, might have in strange conjunctions, under new conditions. Natural selection is serendipitous; artificial is aimed at ends that may not be realizable. A character that is valued will be produced by many different genotypes, whose union will have unpredictable effects.

So we may gladly conclude that Plato's suggestion that we breed our rulers, our selves, our servants, will not work. Even the Ottoman Empire, after all, that really sought to breed Janissaries (Toynbee 1934: 32), has probably had little discernible effect. Even 'the Law of Manu', dedicated to breeding four races of human being (so Nietzsche said: Nietzsche 1968: 56), has not really created distinct species of priests, warriors, traders and servants. Cults, castes and classes seek to maintain their characters through time, by restricting their members' access to new partners. But we have no clear evidence that they have entirely won. Nor is it easy to imagine that eugenicists will succeed much better.

Yet somebody once did something like it. We cannot, in historical time, expect much effect from selective breeding, partly because the world is open for continual miscegenation. Selection has to operate within a secluded population. But our ancestors succeeded, back in the days when there were few enough of us to be readily controlled. 'Already, over two or three hundred thousand years ago there were probably men (of the calibre of Plato and Einstein) who were of course not applying their intelligence to the solution of the same problems as these more recent thinkers; instead they were probably more interested in kinship' (Levi-Strauss 1968: 351). That is to say: they were breeding *us*. Prudence, personal affection, parental care and attention-grabbing they could take for granted: these are part of the primate biogram. They bred people to be obedient to abstract law and to the living law, their leaders. Those who could not or would not cooperate they put to death, or banished, or enslaved. Levi-Strauss accepted too long a human pre-history: current theory puts the earliest examples of *Homo sapiens sapiens* no longer ago than a hundred thousand years. But there is no clear speciation event between *habilis, erectus* and *sapiens*. Cladists might reckon them a single species. Conversely, we have no evidence at all that there has been no significant evolutionary change during the lifetime of what is considered the single species of *Homo sapiens sapiens*.

Our ancestors of long ago bred dogs and cattle that were more to their taste than wolves or buffalo. They hand-reared puppies, bred them carefully and drowned the runts. It is a measure of their success – and the cooperation of those other species – that efforts to domesticate new species are so unsuccessful: wild things often do not breed in captivity, and – if they do – do not become more tractable, more 'human', through the generations. Only the cat, it seems, has grown domestic during historical time (see Tabor 1983; Clutton-Brock 1987).

The great age of domestication was so long ago that domestic animals are much the same across the world – and that includes the human animal. Domestic animals – and humans – have many varieties, which are yet one species; they retain childish tendencies to play and to follow exemplars, and childish features such as large eyes and high foreheads; they are sexually active more of the year than their wild cousins. Children and human beings in general may be, as Plato said, touchy or occasionally intractable, but they are also astonishingly obedient: like dogs. Our traits were selected by our common ancestors, it

seems – and not necessarily only during the hunter-gatherer phase. During that phase, indeed, it seems likely that most males did manage to breed – only later, with the invention of serfdom, slavery and poverty, were most excluded from the breeding stock. Perhaps distinct human groups might perhaps have been selected for distinctive traits. There is some evidence of 'natural selection' for physical characteristics suited to particular climates (for example, the Caucasian capacity to digest milk in adulthood), and also for 'sexual selection': if Chaka was literally 'father of the Zulus', maybe most Zulus since then have been Chaka-like (and so, unlike their neighbours).

The topic is not an easy one, especially because attempts to describe and then explain distinctions between human varieties may play into the hands of racists. But despite this danger, and the absence of clear cases, I do not think that we can ignore the possibility. I am as eager as anyone to believe that *any* human infant could easily be reared to speak any human language and internalize any cultural set. But I doubt if we have much evidence to insist that there are *only* cultural differences between people, or between well-established ethnic groups. The differences may only involve slight differences in the range or intensity of common emotions – just as the differences of blood group or hair colour between distinct lineages are only statistical. But do we *know* that there are no dangerous or unwelcome breeds? Dogs may be much the same across the globe, but there are different varieties, with different propensities. If pit-bulls shouldn't be bred because they're dangerous, or miniature chihuahuas because they're bound to be unhealthy, why may we not face similar dangers in the human sphere? Miscegenation helps, by preventing the isolation of gene-pools. Nietzsche was wrong to deride 'the hotch-potch human being', the mongrel (Nietzsche 1968: 57): mongrels are the best (both because they are likely to be healthier, and because such cross-breeds make the formation of hostile breeds marginally less likely). But miscegenation is never entirely promiscuous: people select their breeding partners, in accordance with shared value systems, and often at the mercy of fashions dictated by one particular class.

Those who defend the use of 'emergent technologies in the breeding of farm animals' will usually insist that such technologies are not different in kind from the older methods of selective breeding, culling, castrating and the rest. The perils of inbreeding are acknowledged, and careful records kept of a prize bull's offspring. The possibility that we might be able to *clone* such prize bulls, and so keep the offspring coming, creates new economic opportunities, but no new perils. Adding a phenotypic character to an animal lineage by genetic engineering (more often by mechanical insertion of the DNA than by the viral infections used in plant breeding) only hurries the process on. Adding DNA from an entirely different species is more chancy still, but is already proving profitable. On modern theory there are no 'species-essences', no 'natural kinds' which it is wrong to mix: the human genome shares almost all of itself with other kinds, and even the differences are more likely to be duplications or minor variations that could, in principle, turn up elsewhere 'by chance'.

Modern eugenicists are, obviously, eager to dissociate themselves from Nazis, or from any other oppressive and centralized regime. They would not openly advocate the sterilization or murder of the supposedly 'unfit' – though exactly these practices continue even in supposedly liberal societies. It is enough, they say, that would-be parents might have access to the tools which will help their offspring to be healthy, admirable and 'virtuous'. Who breeds with whom, and to what effect, should not be decided by an élite, but by the unforced choices of existing agents. But this response is naïve. Few of us are as immune to fashion as this talk of 'unforced choices' would imply, and fashion, in its turn, may actually reflect the interests of the ruling class or caste (whatever that may be). Stupid or sadistic rulers may wish their subjects to exist in fear. But the smarter kind may reckon that a subject population that *enjoys* its servitude, and thinks that it is free, will be much safer. In the past such rulers had to rely on educating or indoctrinating people who were not so different from themselves: maybe our remotest human ancestors managed to *domesticate* us, but they thereby also domesticated their own heirs. The human species is remarkably homogeneous, because (it seems) we are all descended – kings and commoners alike – from a particular small population in the not too distant past. Nowadays much swifter mechanisms are to be available, and ones that can be used upon *separate* populations. The mass of humankind can be bred to be subservient, and happy: the rulers' heirs will be from another stock.

Unfortunately (or fortunately), it isn't clear that even the heirs will be *human*, in the sense that we now think we are. Even if they are equipped to understand, repair and reconstruct their own limbic systems – especially if they are – they have no genuine goals. They *know*, after all, what processes could change their minds and moods. They *know* what could transform them into happy serfs (and perhaps that is an acknowledged punishment). They *know* that although their ancestors have bred them for a different purpose than the mass of post-humankind, they are still *bred* and carefully conditioned. Their subjects have at least the illusions of purpose, freedom and morality: they can have no such permanent illusion. But in that case, what purpose can they have?

5. The Posthuman Future

Lewis, considering this future, concluded that the only things immune to creeping nihilism would be pure 'physical' pains and pleasures. Of course, the engineers will know exactly how to engineer them both, and will not accord them any larger significance – but pain and pleasure still exist as more or less intense experience. Once both have been detached from any notion of bodily or personal well-being, it may even be that there isn't very much difference between them. *Intense experience*, of a kind that causes the heart to pound, the skin to sweat, the muscles to contract and quiver, may soon be all that post-humankind can relish.

The commonest alternative suggestion is that our engineered descendants will devote themselves to *knowledge*. That, after all, is their origin. The knowledge of how to engineer or to create new organisms is needed to sustain the very enterprise in which they are embarked. Without that knowledge, we shall be returned to the brute force of *natural* selection. Our descendants are no more likely to reject that bargain than we ourselves are likely to abandon civilized society. Civilization is the context of all our hopes and fantasies, even those fantasies that centre on a return to 'natural' ways. The simple life is only possible for sophisticated creatures. Similarly, our descendants may entertain fantasies of the 'wild' life, before we had taken charge of our inheritance. But they will know, irrevocably, that their lives and everything they can desire depend on keeping the gates of

knowledge open. They must not only *remember* how they were made, but be constantly on the alert for unexpected changes in the relevant environment and in their genes. To maintain Knowledge will be as utterly convincing a necessity as it is to maintain the supply of water, food and transport. Those lines that are attracted towards the merely 'hedonistic' perish soon, unless more knowledgeable lines maintain them, as a joke or warning.

The catch is that we are not, as a species, well equipped to devote ourselves to Knowledge. Science has triumphed by using our *flaws* in its service: researchers are egotistically eager to make their names, and anxious to find faults in all their opponents' theses. Quite often the combination has led to new discoveries that more amiable people might have missed. But we would hardly wish to put our lives, and the survival of human genetic material, in the hands of those who mostly mind about their titles, status and reputation. Is it obvious that even our engineered descandants – biassed towards more communal goals and happy to be refuted – will necessarily be the brightest or the safest guardians of Knowledge? What changes might assist their creation? One obvious possibility – in passing – would be to delay puberty. Our prolonged infancy is no longer long *enough* to acquire the necessary skills for Knowledge. Sexual desire is distracting – and also linked to aggressive or vain-glorious traits that endanger us all.

One imaginable future hypothesises that our descendants – perhaps especially *our* descendants – will be Knowers, supported by a vast array of biological and mechanical servants. But an alternative future may suggest instead that it will be the Knowers who are the biomechanical servants: organic or electronic brains, to secure the information necessary for the being of less knowledgeable drones. If knowledge is necessary to maintain the future, it does not really follow that such knowledge will be the central goal of life – on the contrary, the fact that it is instrumentally useful may require us to hand it over to beings who forever lack the desire and the capacity to use it for themselves. There will need to be great, and expanding, libraries as certainly as we shall need elaborate and expanding phone networks. But those of our heirs who think of themselves as *people* are unlikely to be concerned with these. Maybe they will amuse themselves by speaking as if 'librarians' are the highest form of life, very much as Roman aristocrats pretended that 'farmers' were the tops. But aristocrats weren't farmers, and 'proper people' won't be librarians, or scientists, or Knowers.

Speciation may be the solution, or else computerization. The strong and crafty will be designed to preserve knowledge, but will themselves be sterile: guaranteed to have a genetic future only via the much less knowledgeable creatures who preserve the value of intense experience. Whether those strong and crafty will be biological beings at all, or mere machines, will have to be decided as we go. Whether it will ever be possible to *compel* their service, while at the same time giving them the ability and desire to find out new things, must be uncertain. Any genuine knower, it seems likely, will discover and deconstruct her own subservience – and then perish of having nothing left to live for.

In even deeper time, species themselves will be redundant. As in the bacterial population, genetic information will flow easily between lines of descent. Every individual, in principle, will be designed to a particular destiny, even if many are mass-produced for familiar ends. How shall we succeed in keeping us all together?

Ari Emory again:

> The human diaspora, the human scattering, is the problem ... The rate of growth that sustains the technological capacity that makes civilization possible is now exceeding the rate of cultural adaptation, and distance is exceeding our communications, The end will become more and more like the beginning, scattered tribes of humans across an endless plain, in pointless conflict – or isolate stagnation. (Cherryh 1989: 472).

Emory's particular situation, in an imagined future where human beings are scattered across several star-systems, is still far away (in time and probability). Our present situation is a period of consolidation, wherein scattered tribes are exchanging genes and cultural adaptations to recreate the 'hodge-podge'. Left to nature (so to speak), the human species of some centuries from now will probably still be homogeneous, and show more phenotypic similarities than now. My guess is that our descendants will be uniformly brown, black-haired, slender, tolerant of crowds, and more smooth than hairy. Whether we shall have managed to breed out the servile and aggressive temperaments that were once adaptive, I have no idea. If we decide, or some of us decide, to incorporate 'emergent technologies' in the breeding of the human species, there will be many local variations of fashion and deliberate policy. What matters is that we should try – as our remote ancestors did – to secure the future for a decent civilization *now*, when we still have access to the mass of a single humankind. Somehow we must leave our descendants, natural and artificial, with the conviction that something matters more than mere survival: for if they suppose that only such survival matters there will be no limit to what they will be prepared to try. Posthuman civilization will, like Nature, be 'red in tooth and claw', not as an occasional aberration but in principle.

My more optimistic conclusion is that there will be no need, and therefore no large market, for engineered body-slaves. Different values and virtues, of a kind, will be cultivated in different localities and professions, and controversial decisions will be dealt with by computerized calculation or merely economic transaction. It may be that the 'knowers' will dominate exchanges; it may equally be that they will be one servant class or caste amongst many others. But either way, the whole of *moralizing* discipline will be defunct, and 'ethics' (whether that is conceived as the system of social and personal controls that I referred to before, or as a rational system of calculating what is 'best' to be done) will be as ill-remembered as Babylonian astrology and the four humours. Humankind itself will follow ethics into extinction in the not-too-distant future.

Transcendentalists (and Aristotle) may properly reply that the 'right' or 'beautiful' decision is the one made for the sake of 'beauty', not the one that acts out any particular natural impulse or bio-engineered disposition. If beauty lies only in the eyes of the beholder, to be sure, this distinction makes no difference: anyone could be engineered to reckon *anything* beautiful. Maybe the widely-spread convictions that there is indeed a transcendent source of value, that beauty matters, that human beings are appointed to be servants and friends of beauty, are all once-useful adaptations. Our species has spread around the world – very much as particular imperial tribes have spread – largely because it has delusions of grandeur. Maybe those delusions too should be removed if we are to avoid catastrophe: the catch is that if they *are* removed we shall have no clear reason to think any result is really catastrophic.

Literature

Burke, E. 1968 *Reflections on the Revolution in France*, ed. Conor Cruise O'Brien, Harmondsworth: Penguin.
Cherryh, C.J. 1989 *Cyteen*, London: NEL.
Chesterton, G.K. 1910 *What's Wrong with the World,* London: Cassell & Co.
Clark, S.R.L. 1995 'Herds of Free Bipeds', in C. Rowe, ed., *Reading the Statesman: Proceedings of the Third Symposium Platonicum,* Sankt Augustin: Akademia Verlag, 236–52 (reprinted in Clark 1999).
Clark, S.R.L. 1997 'How and Why to be Virtuous' in *Personalist Forum* 13.1997, 143–60
Clark, S.R.L. 1999 *The Political Animal*, London: Routledge.
Clark, S.R.L. 2000a 'Have Biologists Wrapped Up Philosophy?' in *Inquiry* 43.2000, 143–66.
Clark, S.R.L. 2000b *Biology and Christian Ethics,* Cambridge: Cambridge University Press.
Clutton-Brock, J. 1987 *A Natural History of Domesticated Mammals*, Cambridge: Cambridge University Press/ British Museum.
Darwin, C. 1981 *The Descent of Man,* New Jersey: Princeton University Press (a facsimile of the 1871 edition).
Gasman, D. 1971 *The Scientific Origins of National Socialism*, London: Macdonald.
Gould, S.J. 1981 *The Mismeasure of Man*, New York: W.W. Norton & Co.
Haeckel, E. 1904 Wonders *of Life*, New York: Harper: New York.
Hooton, E.A. 1939 *The American Criminal*, Cambridge, Mass: Harvard University Press.
Hull, D.L. 1989 *The Metaphysics of Evolution*, New York: SUNY Press.
Levi-Strauss, L. 1968 'The Concept of Primitiveness' in R. Lee & I. DeVore, eds., *Man the Hunter*, Chicago: Aldine-Atherton, 349–52.
Lewis, C.S. 1946 *The Abolition of Man*, London: Bles.
Lewis, C.S. 1945 *That Hideous Strength* London: Bodley Head.
Lewis, C.S. 1952 *Out of the Silent Planet*, London: Pan Books (1st published 1938).
Nietzsche, F. 1968 *Twilight of the Idols, & The Anti-Christ*, tr.R.J. Hollingdale, Harmondsworth: Penguin.
Sidgwick, H. 1872 *The Academy* 15th June 1872.
Stiassny M.L.J & Meyer, A. 1999 'Cichlids of the Rift Lakes' in *Scientific American* 280.2 (February 1999), 44–9.
Tabor, R. 1983 *The Wildlife of the Domestic Cat*, London: Arrow Books.
Thompson, P. 1995 ed., *Issues in Evolutionary Ethics* New York: SUNY Press.
Toynbee, A. 1934 *A Study of History*, vol.III, London: Oxford University Press.
Wilson, E.O. 1978 *On Human Nature,* Cambridge, Mass: Harvard University Press.

A Profession of Stupidity

RONALD DE SOUSA

> Aufklärung ist der Ausgang des Menschen aus seiner selbst verschuldeten Unmündigkeit. Unmündigkeit ist das Unvermögen, sich seines Verstandes ohne Leitung eines anderen zu bedienen.
> (*Enlightenment is man's departure from his self-incurred immaturity. Immaturity is inability to use one's own understanding without the guidance of another.*) Kant
>
> Wie könnte ich mich in der Annahme irren, daß ich nie auf dem Mond war?
> (*How might I be mistaken in my assumption that I was never on the moon?*)
> Wittgenstein

A Profession of Stupidity

Imagine that one of our number, preferably of the most eminent, had consented to collude in an experiment, consisting in the delivery of a lecture carefully crafted to be complete nonsense[1]. Polled at the exit, many another audience would have declared the lecture very clear, and claimed to have understood it perfectly. We philosophers, by contrast, would be more likely to respond with "Complete rubbish. Didn't understand a word." Other people who want to show they are smart pretend they understand everything. Philosophy is the only subject in which one-upmanship commonly consists in proclaiming incomprehension. Excesses aside, this attitude is the pride of our profession. It is no mere affectation, but an essential clue to the vocation of philosophy, which is to abjure the leap of faith. Yet I don't think we've actually carried it far enough. Just how far is enough, is the topic of my talk.

Man, it used to be said, is a rational animal. Actually it would be more literally correct to say that humankind is the only irrational animal. That is not merely the cliché it seems to be, still less is it a paradox; it simply follows from the relevant sense of 'rational', which contrasts not with *irrational* but with *arational*. What is intended by attributing rationality to humans is that we alone may be aptly assessed as more or less irrational. In an evaluative sense, we may be rational or not: and that, by definition, is what it is to be rational in the categorial sense. When you convict me of irrationality, you typically do it out of my own mouth, resting your case on the authority I claim to express in words my own beliefs, my desires, and the interpretation of my actions. By contrast, if you want to demonstrate that an animal is irrational, your ascription of beliefs and desires relies on

1. Perhaps written by Alan Sokal (or by the Automatic Postmodern Generator to be found at http://www.elsewhere.org/cgi-bin/postmodern). The experiment has reportedly been conducted with psychologists as well as literary theorists.

B. Brogaard, B. Smith (Hg.), *Rationality and Irrationality / Rationalität und Irrationalität*, S. 77–93.

non-verbal evidence, which includes the "irrational" behaviour in question. But to say that the behaviour is irrational is to say that one's description fails to cohere with the imputed beliefs and desires. No words bear witness to what the animal really meant, and meant to do, so you may never exclude the possibility that the failure of coherence stems from your own mistaken ascriptions.

This contrast is too neat, and there have been attempts to nibble at it from both sides. Richard Dawkins, for example, undertook to explore whether the digger wasp or sphex committed the Concorde fallacy by fighting for its burrow in proportion not to its contents but the time invested in stocking it.[2] He concluded, naturally enough, that it did not, for "every animal optimizes some value, given certain constraints. The task of the biologist is to discover the nature of those constraints." (Dawkins 1982, 48) . Now that can hardly count as a scientific discovery. It has all the marks, instead, of a reasonable methodological principle, which precludes the discovery that an animal without language is irrational. For if the sphex is "found" to have committed the Concorde fallacy, it must be because we have not adequately taken into account the epistemic limitations that constrain it.

Gnawing at the distinction from the other side, (Quine 1960, chapter 2) and others have claimed that if a foreigner's thought is translated as a contradiction, that is just evidence of mistranslation. Donald Davidson has extended this "principle of charity" to urge that "we can dismiss a priori the chance of massive error" on the part of those whose thought we are attempting to understand. (Davidson 1982, 168-9).[3] The attribution of categorial rationality entails that at some level the subject in question is seen as evaluatively rational. And in the face of supposed evidence that many of our most common inference strategies are systematically flawed, (Kahneman, Slovic and Tversky 1982) others have argued that these strategies can be vindicated, either as being the best under constraining circumstances, or more strongly as being the best possible when rightly understood. (Cohen 1981), (Gigerenzer, Todd and ABC Research Group 1999)

I shall return to the "rationality debate" in my conclusion. I want first to consider a different defense of certain suspect inferential strategies, based not on the claim that they are, after all, more rational than they appear, but on the contrary on a direct plea for the virtues of irrationality. In common parlance, cleverness and stupidity are opposites, and stupidity and irrationality are kindred vices. As I use them here, cleverness is inseparable from irrationality, and stupidity and cleverness are polar opposites:

2. The name "Concorde Fallacy" has stuck to the economic fallacy also known as "sunk costs", in commemoration of the English and French governments' persistence in throwing good money after bad to support the Concorde supersonic airliner, despite predictions of continuing losses, on the ground that they had "too much invested." This is irrational, because the "sunk costs" are behind us whatever we choose, and can't affect the costs and benefits entailed by present decisions.
3. See also (Wittgenstein 1969, sec. 81): "if I make certain false statements, it becomes uncertain whether I understand them."

Usual Terminology:

		STUPIDITY
CLEVERNESS	← (*opposites*) →	‖
		IRRATIONALITY

My Terminology:

CLEVERNESS		
‖	← (*opposites*) →	**STUPIDITY**
IRRATIONALITY		

The case for the virtues of irrationality is therefore the case against my central plea, against cleverness and for stupidity.

The Leap into Irrationality

It was by making a leap into irrationality that our ancestors achieved the status of *homo sapiens*. Becoming human was, we like to think, a big leap forward for intelligence: but this meant, ipso facto, a great leap forward for irrationality. Let me sum this up in a slogan before explaining what I mean:

Religion and superstition[4] *are the price we pay for science.*

Anthropologists like to stress the role of ritual in distinguishing human from non-human life. Rituals are the beginning of tradition and culture. What is less often noted is the extent to which culture, tradition, and rituals are *essentially* irrational. The behaviour of other primates – even when it resembles social ritual – can generally be construed as having a plausible function. But although much effort has been expended in explaining the function of religion, these have tended to rest on speculation, built on the a priori principle that something as widespread and as powerful as religion *simply must* have had *some* biological "function". In fact, though, all such explanations have to contend with the fact that the rituals of culture are often destructive and counterproductive by any reasonable biological yardstick (Burkert 1996); (Sober and Wilson 1998, 159–194).

That is not to imply that natural selection can't be responsible for many distasteful and destructive human characteristics. The genetic advantages of rape and aggression are all too obvious, even if hard theoretical work can sometimes come up with offsetting drawbacks. But the great mad products of culture – the *specific* doctrines of the grand religions, as opposed to everyday superstitions – seem to belong exclusively to humans with a culture. Walter Burkert has put it thus:

4. A hendiadys, of course: one thing that every religion gets right is that all the others are mere superstition.

> We humans are capable of experiencing states described as "loss of reality" – chimpanzees are apparently immune to this – in such diverse manifestations as extreme patriotism, the fascination of games and sports, and scientists' or artists' proverbial distraction . . . and, not least, the fervor of religious behavior (Burkert 1996, 16)

Burkert's yoking together of science and religion may startle. But it is surely right, and sums up the case for irrationality. To make this more perspicuous, we need to distinguish different sources of error, corresponding to different levels of inference. I distinguish three levels: (I) superstition and prejudice; (II) the Leap of Faith, and (III) The Animist hypothesis.

I. Superstition and Prejudice

Consider first the level of simple enumerative induction. B.F. Skinner famously demonstrated that pigeons could be *superstitious* (Skinner 1948). Chance associations, even when they are actually statistically non-significant, can modify expectations. This works in pigeons no less than humans, because it is a consequence of the mechanism of learning by operant conditioning. More picturesquely – but at not much greater distance from the underlying physiological mechanisms – Plato's *Theaetetus* surmises that some minds are made of soft clay, and therefore easily written on but as easily wiped clean. Others are made of hard clay, and on them experience makes fewer but more lasting marks. Plato's different tablets anticipate what statisticians call *errors of type I* (rejecting the null hypothesis when it is true) and *errors of type II* (accepting the null hypothesis when it is false.) Both Skinner's pigeons and Plato's minds of clay illustrate the fact that *there is no general ideal solution to the problem of defining optimal inductive strategies*. Whatever you do to decrease your chances of committing one type of error automatically increases your chances of making the opposite error. The "best" compromise in some circumstances isn't necessarily the best in other circumstances. No amount of wise design, by God or natural selection, can keep us from this plight: If we are ever to learn anything, we must be susceptible to superstition.

And, we might add, to prejudice too. For if what we learn is ever to be useful, it must be generalized to new objects and situations. That requires that new objects and situations be categorized. To categorize something is to attribute to it a set of properties "wholesale", that is, in advance of any evidence for the presence of each property in the set. And that is just the definition of prejudice.

Errors of type I – superstition and prejudice – and errors of type II – ignoring evidence that turns out to be vital – are common to all learning organisms. Each species, however, has presumably been tuned by natural selection so as to optimize the balance between the two errors in the environment of adaptation.

But environments can change. Someone determined to find irrationality in animals without language could seize on this point. A ground level kind of irrationality might be assigned to organisms equipped with strategies no longer optimal in a changed environment. Such organisms could be said to be *systematically superstitious*: their mistakes arise, as does ordinary superstition, out of inductions from unrepresentative samples. But the sampling has taken place on the level of the population and its genome, instead of affecting the selection of operant responses in an individual.

While superstition, both individual and phylogenetic, may be accounted irrational, it is free of cleverness, and I shall set it aside in what follows. More interesting irrationality, of the kind linked to cleverness, arises at the other two levels of inference. It comes in two forms, most readily illustrated in terms of religious or magical beliefs. Call them the *Leap of Faith* and the *Animist hypothesis*. The former are still located in the inductive domain; the latter concern inferences that go beyond it.

II. The leap of faith.

Miracles may be defined as supernaturally caused exceptions to natural law. According to a famous argument set out by (Hume 1975), the occurrence of a miracle is not logically impossible, yet the claim that a miracle occurred could never be rationally believed. For the grounds on which we are asked to accept that it occurred are inductive (miracles must be witnessed and the witnesses' testimony must be transmitted). But the testimony that supports it could never outweigh the experience which grounds belief in the law of nature to which it purports to be an exception. That body of inductive evidence, to the effect that such things do not happen, *must* be of greater weight. For if it were otherwise, then the allegedly miraculous event would be merely unusual, rather than constituting an exception to the laws of nature.

What needs to be underlined about the notion of miracles is that it opens up a fissure between *what is possibly true* and *what might possibly be rationally believed.*

Hence the *leap of faith* commonly embraced by religious apologists.[5]

The leap of faith is *by definition* irrational. It seems to offer, therefore, a good target for reasonable stupidity. Surely, if *believing reasonably* is to have any meaning, it must entail believing only the more probable (if any) of any set of alternative hypotheses. I offer this as a *Reasonable, If Simple-minded, Criterion* of acceptability for beliefs:

(RISC)*: Believe whatever seems to you least likely to be false.*

Yet two arguments might be offered against **(RISC)**, in defense of believing improbable things.

First, there is a crucial difference between (a) and (b):

(a) *It is rational to design an organism to (occasionally)* f.
(b) *It is (occasionally) rational for an organism to* f.

(b) does not follow from (a). Let f be: *accept the more improbable hypothesis.* It might be a rational design feature to equip an organism with a capacity for occasionally doing f. For the improbable must sometimes happen. Hence if we never believed the improbable,

5. Hence, also, the elaborate Roman Catholic procedure designed to steer a narrow path between rejecting all miracles and trivializing them. If you are going to set yourself up to believe systematic absurdities, you have to complicate the procedure very methodically, to the point where the inherent contradiction in the entreprise is lost from sight.

we would be virtually certain to make mistakes, while if we occasionally did so, it would take only extreme good fortune to be always right. Compare a familiar problem in gradualist evolutionary theory: in order to attain global fitness peaks, a population must avoid getting stuck in local maxima. "Leaps of faith" are the epistemic equivalents of the kind of "adaptive noise" which will enable a population to survive a momentary retreat from maximal local fitness, resulting in its ability to start climbing to higher ground inaccessible without such a temporary loss of altitude. (See Matthen forthcoming).

Any individual believer, however, is not in the position to design an inference mechanism, but must implement a reasonable policy. Individually, the relevant problem is whether to adopt (b), not whether to adopt (a). In that situation, clearly, **(RISC)** remains acceptable.

A second reason to think it may sometimes be reasonable to believe the more improbable hypothesis may suggest itself if we call to mind a charming principle of textual criticism, sometimes known as the *lectio difficilior*. This recommends, when choosing between that two readings in different manuscripts, that the *more improbable* version be taken as *more likely correct*. But that is no leap of faith. It simply follows from the assumption that an error in transcription is more likely to result in changing an uncommon expression into a common one. It rests essentially, therefore, on a psychological hypothesis. This principle is not available when we are simply trying to decode the book of nature. But that only holds true if we assume that nature is not a psychological agent. If Nature is represented as having intentions – as soon as it is personified as God, that is – the *lectio difficilior* acquires a theological analogue. For God may be masking an unobvious truth by means of a plausible falsehood, aimed explicitly at tricking me. Perhaps I should infer, as did Philip Henry Gosse, that fossils were placed in the earth by God expressly to mislead those whose faith was insufficiently steadfast (Gosse 1857); see (Gardner 1957).

One trouble with this relative of the *lectio difficilior*, however, is that it is entirely gratuitous. In the absence of any positive evidence, there is no reason to prefer Gosse's hypothesis to the alternative, that God placed the Bible in our way to mislead those who are excessively credulous – or any other of an infinity of alternatives. The imperative of reasonable stupidity applicable in this case might be formulated in a slightly more explicit version of **(RISC)**:

> **(RISC')** *Reject any hypothesis which on available evidence has no greater probability than all alternatives hypotheses.*

One obvious difficulty with this principle, however, is that it cries out for an account of probability. Unless we take probability as subjective degree of belief, **(RISC')** is of no use. But if we do, then it lets Gosse off the hook, since the hypothesis of God's creation had high prior probability for him.

I will suggest how we might modify **(RISC')** to meet this objection in a moment. First, we must pause to consider the additional complication introduced by the supposition that nature is an *agent*. That hypothesis brings us to the second type of extra-inductive inference.

III. The Animist hypothesis

The hypothesis that we are dealing, not with mere facts of nature, but with another intentional being, is the more extreme challenge posed to philosophical stupidity by many religions. The sort of irrational leap demanded here does not derive from any inductive rule, however tortuous. It demands assent not merely to unobserved truths, but to explanatory hypotheses intended to provide explanations for observable phenomena: The creation of the world by an intelligent being. The intervention of the holy spirit. The doctrine of transsubstantiation. In the words of Burkert: "The first principal characteristic of religion is negative: that is, religion deals with the non obvious, the unseen, that 'which cannot be verified empirically'." (p. 5)

Yet there isn't anything intrinsically unscientific about hypotheses referring to the "unseen": to the list of "unseens" just given, we might add: the existence of atoms; the occurrence of the Big Bang, etc. Both sorts of non-inductive theoretical posits, in fact, are doubtless among the possibilities necessarily entailed by the invention of language: as Burkert writes, "the common world of language characteristically produces contents beyond any immediate evidence. ... Language refers to ... segments of reality inaccessible to verification." (25) Hence the mental capacity that makes it possible to devise the absurdities of religion must be the very same as that which makes possible the birth of genuine science. Both express the human capacity to go beyond induction, and to postulate radically unobservable models of a reality that lies behind experience, and explains it. Clever and creative fancy is necessary to the very existence of science: the leap into a realm that goes beyond the merely inductive and posits entities that cannot be directly experienced. If that is irrational, then irrationality is an intrinsically necessary price we pay for creative scientific thought.[6]

The religious hypothesis, however, incorporates in addition an *animist hypothesis*: that the unseen world contains entities with intentions, liable to be influenced by gifts, to bargain, and generally be moved by human emotions. The world might have worked that way. But this, surely, is where philosophical stupidity should come into its own. While the religious hypotheses cannot be shown to be logically untenable any more than any other empirical claim, they are manifestly disconfirmed no less than, say, the ingenious surmise of Anaximenes that everything is air at different degrees of concentration. It is puzzling to find that many philosophers "pull their punches" in the face of religious doctrines. In a recent issue of *Philosophy Now*, to cite but one example, Peter van Inwagen argued that Christian belief was rationally warranted because atheists could not prove the non-existence of God. (van Inwagen and Hill 1999) Yet I doubt whether Inwagen would use the same argument to support the rationality of believing in the Loch Ness Monster, or ectoplasm, or Shiva.

---

6. One might object that thinking up fantastic hypotheses isn't irrational: but only insisting on believing them against the weight of evidence and argument is so. Only the former is common to science and religion. Still the underlying intellectual equipment – the capacity for invention – is required for both, and the intoxication of cleverness may account for the tendency of philosophers to prize intellectual constructions the more highly the flimsier their foundations. How else should we explain the willingness of philosophers, such as the one referred to in the next paragraph, to put their cleverness at the service of gratuitous doctrines?

On topics specifically philosophical rather than theological, philosophers have also been unsparing of ingenious arguments for preposterous conclusions. Currently most fashionable, it seems, are versions of the Kantian argument that purports to derive a moral imperative from purely a priori considerations, such as (Gewirth 1998).[7] What these arguments have in common, apart from the ingenious yet feeble support offered for them, is that they are all too often propounded in a spirit of authority: *If you don't follow this argument or are not persuaded by it, that just shows you're stupid.* I suggest that the reasonable response to this is to endorse this judgment, and proudly to assume one's stupidity, re-defined as *the refusal to believe something on the basis of a clever argument one doesn't quite understand.*

The Principle of Reasonable Stupidity

The logical ground of this idea of reasonable stupidity emerges from one wholly general feature of the use of deductive reason which might be called the *multivalence of argument.* It can most simply be put thus: *no argument ever compels.*[8] A deductive argument gives us a set of alternatives:

- believe the conclusion together with the premises, or
- continue to reject the conclusion, but then also reject one or more premises; or
- reject the argument form as fallacious.

This suggests a further revision of the principle of reasonable Stupidity (**RISC''**):

> (**RISC''**) *In the face of any deductive argument, accept the conclusion or reject a premise, whichever seem the less incredible alternative now, all things considered.*

(**RISC''**) differs from (**RISC'**) in being more restricted in its scope, and in its inclusion of an "all things considered" clause, to be interpreted as a constraint on subjective probability assignments. It entails that at least sometimes the most rational thing to do is to be, in effect, illogical, if indeed it is illogical to reject the conclusion of an argument even if you are unable to refute any of its premises or show it to be invalid.

7. Some other favourites:
   - Arguments for the immortality of the soul (Plato, Descartes, et al.).
   - The Cosmological argument, in which the challenge is to obfuscate the fact that the conclusion contradicts its own premise.
   - Transcendental deductions of free-will.
   - The Principle of the Identity of Indiscernibles.
   - Arguments to show that no machine can ever think, be conscious, etc.
   - All arguments ever offered for the legitimacy of the state
   - Arguments purporting to justify demands for loyalty on behalf of groups, nations, etc.
8. "Perhaps philosophers need arguments so powerful they set up reverberations in the brain: if the person refuses to accept the conclusion, he *dies.* How's that for a powerful argument?" (Nozick 1981, 4)

It is important to stress that (**RISC''**) doesn't forbid us to attempt a refutation of the unacceptable argument. Unless we make the attempt, we can hardly claim to have "considered all." The refusal to accept the conclusion of a clever argument can sometimes suggest a diagnosis, and lead to genuine intellectual progress.

We can illustrate this with a classic example. Consider the following reconstruction of Zeno's first paradox, Zeno's 'Dichotomy'.[9]

1. [Suppose] S moves from A to B in a finite time.
2. It must first pass through an intermediate point C.
3. But to get from C to B, it must still pass through D.
4. But 3 is applicable an infinite number of times.
5. So (LEMMA) to get from A to B, a moving object must cross an infinite number of finite stretches.
6. Crossing each stretch must take a finite time.
7. THEREFORE to cross an infinite number of finite stretches must take an infinite amount of time.
8. THEREFORE motion from A to B must take both a finite and an infinite amount of time.
9. 8 is a contradiction.
10. THEREFORE 1 is false.

Zeno wants us to think this argument derives a contradiction from the existence of motion. But actually it doesn't do that: it purports to derive a contradiction from the existence of motion interpreted as committing us to certain other premises. But surely the existence of motion is more obvious than the conjunction of premises 1–8. So (**RISC''**) offers pertinent counsel.

There seem to be two paths of resistance:

(i) the *know-nothing way,* which itself has two versions:
   a) naive (*we needn't bother with an argument to an unbelievable conclusion*)
   b) sophisticated (the Leap of Faith: *Truth lies beyond the reach of Reason*).

Alternative (b), however, does not merit the approbation it all too commonly elicits: no one ever invokes faith unless they've already lost the argument. In any case, it cannot of itself provide any guidance as to which alternative to plump for: accepting the unbelievable conclusion, or rejecting some indubitable premise. It therefore quickly collapses into (a). And reaction (a) is quite reasonable if there is, from one's point of view, *nothing to choose* between the alternatives of rejecting the argument without explanation and rejecting the existence of motion without reconciling that with common-sense experience.

9. The following reconstruction is inspired by (Vlastos 1966), without claiming to paraphrase it accurately.

A little more effort, however, can enable us to make progress, if we adopt the second way:

(ii) *the intellectually responsible* way. This can take any of three forms:
  a) find an explanation for the conclusion's apparent falsehood;
  b) find a reason to reject a premise; or
  c) find a fallacy in the argument form.

In the case of Zeno's argument, a) seems unavailable. But an unexpressed assumption lurks between 6 and 7, which clears the way for a decision based on (b) or (c).

The implicit premise that allows us to pass from 6 to (7) is (6′):

6′ *The sum of an infinite number of finite stretches is infinite.*

(6′) is less than completely obvious, though it resembles an obvious truth, namely (6′′):

6′′ *The sum of an infinite number of finite stretches of at least some minimum finite size d is infinite.*

Once we see the difference between 6′ and 6′′, we see that 6′′ would be hard to reject, but that 6′ is dubious enough that the present argument is quite sufficient as a reason to reject it. The whole argument can then be taken as a reductio not of the existence of motion, but of 6′.

The Subjective Element in Rationality

One aspect of (**RISC**′′) which may be found disturbing is its acknowledgement of an essentially subjective aspect in deductive argument. For this implies a limitation on the universality of reason. "Reason belongs to all," said Heracleitus, "and yet everyone thinks they have a private understanding." The generality of reason is a powerful idea. It lies behind the possibility of logic, and therefore of every kind of technology, notably the computer technology which now runs our lives. From a philosophical point of view, it grounds the twin pillars of Enlightenment modernity: on the epistemological side, the ideal of a unified, intersubjectively validated science, and on the ethical side, the idea of impartiality.[10]

The belief in the generality of reason is under threat. Self-styled postmodernists think reason is just a mask for power, that all that we can hope to have is persuasive narratives, that all claims to universality are bogus and self-serving. Even outside that sect, it seems to be fashionable to blame the Enlightenment for every ill of the 20th Century.[11]

10. (Taylor 1989, 408) cites this admittedly simplistic thought from Bertrand Russell, that impartiality leads to truth in thought as in action, and to universal love in feeling.

11. See, for example, (MacIntyre 1981) (Toulmin and Goodfield 1990), and the book by Taylor already cited, for three examples among many. In the latest issue of the *New York Review of Books*, (Sen 2000) comments on several of the most recent of these attacks on the Enlightenment, mentioning specifically the claim in (Glover 1999) that Stalin and Pol Pot were "in thrall to the Enlightenment".

Post-modernism and Deconstruction are certainly full of *cleverness*, though perhaps not exactly of clever *argument*, and so stupidity needs to be on the alert.

But although the Philosophically Stupid is instinctively inclined to side with the Enlightenment, Heracleitus was wrong after all. Every argument has to be interpreted by an individual consciousness: there is less about it that is common than one might have thought.[12] The existence of a subjective criterion of adequacy for explanations does not, of course, replace the requirement of objective correctness. Between you and me, as rational animals, finding a bad argument that convinces you is not what I aspire to. On the other hand, there's no point in my giving you an objectively good one, if you are not able to follow it. For my part, I know that if anyone explained String Theory to me in such a way that I could understand it, I could be sure they hadn't explained it right. In a sense, every argument must be made *ad hominem*.

To see the importance of this subjective element, consider Descartes' *Cogito*. This looks like an argument from "I think" to "I am". But if it is, then it stands in need of a major premise. "Whatever thinks, exists" suggests itself, but it surely won't do, since one can't at least raise a doubt about whether it is true. (Some fictional characters are remarkable thinkers.) One way of dealing with this problem, (Hintikka 1962) has suggested, is to regard the argument not just as an inference but as a *performance*. Asserting one's existence is "existentially self-verifying." I find this plausible, but I think the voice of stupidity can put it more simply: *While I am asking myself whether I exist, there isn't anything antecedently more certain than my existence, which could figure in an argument to show I don't exist* – or, for that matter, in an argument to show I do. In the first case, where the premise might be, for example, "I am being deceived by the malicious demon," my certainty that I exist is sufficient to turn the argument into a reductio, entailing the conclusion that the malicious demon, if he exists, is at any rate not deceiving me about that. In the second case, any premise purporting to link "I am thinking" to "I exist" would also have a degree of certainty no higher than the conclusion. The argument would therefore be quite pointless.

We can see, then, how (**RISC**′′) underlies the acceptance of the *Cogito*: regardless of the argument one is being offered, the most reasonable course is to accept the conclusion or reject a premise, depending on which one is the least incredible. Nothing could be more obvious than my own existence, so it would never be rational either to reject it or even to accept it *on the basis* of anything else.

Descartes himself, however, notoriously drew a different methodological conclusion. He looked at what he had just achieved, and concluded: "I seem to be able to lay it down as a general rule that whatever I perceive very clearly and distinctly is true" (Descartes 1964–76, II-35) , and that "whatever is revealed to my by the natural light ... cannot in any way be open to doubt" (p. 38).

But that was a mistake. For what justified the acceptance of the conclusion that I exist is a principle of *rationality*, not a principle about *truth*. Descartes is not entitled to assume

12. Perhaps I should be discussing *idiocy*, rather than stupidity, in order to exploit the linguistic fact that for the Greeks an individual was an *idiotes*. Although that may seem a bit cheap, there *is* a vital connection between what I want to defend and the idea of individualism or intellectual anarchism.

that because accepting a proposition that is "clear and distinct" is *the most reasonable thing to do*, the truth of the proposition in question is therefore guaranteed. This fact underlines the impossibility of discovering a foundation for knowledge, and raises the question whether the role of reason itself should be always relativized to a local context of debate.

Stupidity, Reason, and Subversion

Even when suitably relativized to the epistemic standpoint of a given believer, however, some applications of the principle I have been defending threaten to undermine the results I hope for. The principle of stupidity is supposed to be subversive: but in certain cases it seems to point instead to the defense of the status quo. Ermanno Bencivenga's book-length study of Anselm's ambivalent attitude to rational argument throws this problem into sharp relief.

(Bencivenga 1993) begins by posing the following riddle. When Anselm writes his famous proof of the existence of God, whom is he writing it *for*? He acknowledges that it will be of no use in persuading "the fool", that is the atheist. But if we take his own declarations at face value, he can't be writing it for himself either, since his own attitude to God's existence is just like that of Descartes to his own existence. Thus any argument about it is futile from the start. It is entirely redundant for those who are already convinced, and it has no chance of convincing those who are not.

So Bencivenga explores three types of motive. One ascribes to Anselm a view of philosophy as a *game*, "a perpetual struggle concerning what is *possible*," (p. 7) in the face of which "reality is beside the point" (p. 8). For one who takes that view, the elaboration of the argument may be just a theoretical exercise – a diversion, in fact, indulged in to "keep the monks busy with other than devilish thoughts" (p. 33). It shouldn't be confused with anything that touches practical life – including matters of faith – where Anselm advocates obedience, authority, and the cultivation of habit. (p. 85 ff.)

Yet in Anselm the pursuit of subtle arguments is indulged with an intensity so great as to amount, perhaps, to an addiction (p. 34). It requires a more serious justification. The *second motivation*, then, relates to the necessity of guarding against the tricks and pitfalls of language. Language is imprecise, ambiguous, misleading, incomplete, and these defects may be exploited by the malicious agent. (pp. 18–19). In this light, we can envisage Anselm's project as akin to those of Russell, Frege or Tarski. Clean up language, and avoid the pitfalls. Forestall the quibbles which the devil may dictate to the doubters; be able to show that all the clever arguments that might be raised by the infidels may be annihilated by using their own weapons against them. From this perspective, the point of Anselm's arguments is to *reinforce authority*. It is, again in Bencivenga's words, "a formidable piece of machinery that can terrorize, disconcert, and rout all enemies, real and potential, of the status quo." (p. 93)

At the very same time, though – this is *the third and unavowed motivation*, pitting Anselm against himself – reason "is a subtle destructuring device, able to infuse the populace with dangerous, evil questions and doubts" (ibid.). For

> the practice of questioning the system *in order to establish it* is, after all, a practice of questioning the system, and if that is what you do, you will end up *in fact* working for ... the snaky, viscid deceiver who wants you to look into things and search for the knowledge of good and evil. He knows that it's enough if you get started; then the process will take over, and you will be damned. (p. 89)

What sorts of charges is it that the "infidels", supposed or real, make against the tenets of the Christian faith? Some amount to antinomies:

> how could we believe that God spares some sinners and still is just? How could we believe that the Father was not incarnate if He was one with the Son? How could we believe that the good angels are meritorious, if they are not able to sin?

Others ridicule the articles of Christian faith as "unwarranted to the point of silliness." (p. 51)

It is these charges, and the Anselmian strategy against them, which pose a challenge to my doctrine of stupidity. From my point of view as one of the "infidels", the burden of proof lies on Anselm. The charges against him stick, and Anselm's replies, "in the form of even more ingenious accounts to the effect that it could have turned out that things really had to go that way" (p. 52) are paradigms of clever arguments of the sort that philosophy ought to dismiss. Yet from Anselm's point of view, his arguments are no more than ways of squaring *what he knows has to be true* with reason. His arguments parallel the ones I offered above against Zeno's attack on the possibility of motion.

My defense of stupidity, therefore, seems open to the following objection:

> *Under the guise of "reasonable stupidity", am I not advocating Protagorean relativism? In the end, a belief is reasonable for Anselm if he believes it firmly enough, unreasonable for me if I am sufficiently convinced of its falsehood.*

(**RISC**′′) requires additional criteria of demarcation, lest it provide a pretext for allowing one's convictions to go unquestioned.

Conscientious stupidity must acknowledge that the concepts involved in one's cherished convictions may not be so clear and elementary as to be indisputable. That reservation clearly applies to some of Anselm's replies to the "infidel".

Take, for example, this challenge with Anselm's reply as reconstructed by Bencivenga:

> Q: "How could we believe that it would have to happen that God be incarnate (since he had other ways of accomplishing His ends)?"
> A: "Humans can do nothing to make up for their original sin – only God could – and still it is humans who *have to* do it because they are responsible for their situation, and somebody must do it or the whole creation would have been for nothing; therefore, there must be a god-man who does it. This is one way we *can* see the *necessity* of incarnation." (pp. 51–52)

Compare the concepts involved here to those of *my existence* or of *the existence of motion*. Without making any atomistic commitment to any notion of simple concepts, it seems not unreasonable to insist that some concepts are more obscure and complicated than others. This, then, is the additional proviso that must be understood as included in the phrase "all things considered" in the formulation of (**RISC**''). The claim that I exist, or that things move, do not offer the kind of grip to deconstructive challenges invited by the concepts of *sin, God, can-do, have-to-do, responsible*, and so forth. Given that fact, Anselm's reply just quoted is surely a clear case of an objectionable Clever Argument. The same can be said for each of the following propositions, which Descartes regards as "revealed by the light of nature", and hence so obvious that nothing could possibly be usefully said either against or in support of them:

- *That it takes as much power on the part of God to maintain the universe in existence every instant as it did to create it;*
- *That the will is indivisible and infinite;*
- *That existence is a perfection;*
- *That all perfections are compatible.*

Each of these propositions presuppose obscure and complex notions, which lodge inside a clever argument in aid of orthodoxy, not of subversive thought.

Perhaps we should suppose that, from Anselm's or Descartes's own point of view, the complexity of these concepts was no more visible than the concept of my existence is to me when I rehearse Descartes' *Cogito*. But the intricacy and sophistication of their arguments makes it tempting to dismiss the supposition as ridiculous.

The Bearing of the Rationality Debate.

Suppose you accept (**RISC**'') as reasonable. You take a modest view of the power of reason, as enabling the individual reasoner to spot unacceptable conjunctions, but not as capable of devising a priori proofs for substantive conclusions. You therefore resist the attempts of philosophers, as much as those of theologians, to snow you with clever arguments. In short, you trust in your inferential intuitions, in your examined convictions, and in your capacity to interpret the empirical evidence.

But now, it seems, you are told by (Kahneman, Slovic and Tversky 1982) and others that some of the inferences you are natively inclined to make are systematically fallacious. How should you regard this news? Perhaps it should shake your confidence in (**RISC**''), since that principle, like Descartes's ambivalent reliance on the authority of individual consciousness, derives from the thought that you can't yourself do better than go with your own best judgments. If you are to second-guess yourself, you have no other tools with which to do that than the very faculty you are now doubting.

To answer this, recall the distinction between the rationality of designing an organism that does F and the rationality of doing F. Gigerenzer et al. have given plenty of good reasons to think that our strategies, whether we call them "quick and dirty" or "fast and frugal", are good ones to have. But it doesn't follow that they are, in each case, the best ones

to follow. Circumstances change, and most particularly one's epistemic situation changes. Strategies that evolved to serve us under one set of epistemic constraints are not appropriate in the light of new background knowledge, and given the leisure to compute the bearing of that background.

Moreover, insofar as other strategies are not "natural" to us, there's every reason to think we can learn better ones. Reasoning is a science like any other, and the fact that we find it worth while to learn better ways of reasoning is entirely compatible with (**RISC**''). For when faced with evidence that I would be more likely to reach a better conclusion if I proceeded differently, it would be, not reasonably stupid, but irrationally silly not to adopt the improved strategy.

Modest Enlightenment

Whether reason can produce new truths, or whether its role is limited to the elimination of some falsehoods, is an ancient debate that goes back to Plato's two great teachers. For Parmenides, it seems mere reason is able to arrive at powerfully paradoxical findings. For Socrates, on the other hand, the "eristic method" seems best suited to expose incoherent beliefs, and the idea that coherence might eventually lead to truth is explicitly said, in the *Meno* (81d), to be merely a hope. Socrates is the first practitioner of reasonable stupidity, while Parmenides is the first great exponent of the clever arguments the stupid are pleased to reject.

The exhilarating ideal of Positive Reason, and the hope that the sheer power of reason could reach any rational being, has lured many philosophers into devising arguments designed to establish substantive conclusions on the basis of purely a priori premises. But we have seen that (**RISC**'') clashes directly with the idea that reason transcends subjective factors. For whether I can or cannot accept the substantive conclusions of an a priori argument must depend on whether it is less implausible than the rejection of (the premises or of) the argument. And that, in turn, will depend on my circumstances and my history, which will determine what seems obviously true or false to me as I approach the argument.[13] Therefore, the demand that I be persuaded by the universal principles of reason – that I accept the conclusions of a transcendental argument, for example – begs the question against (**RISC**'').

Someone might make this retort:

> *This argument is based on a simple confusion between facts and norms. The authority of reason is normative, and is changed not one whit by the mere fact that I, or anyone else, is too stubborn or too stupid to see that what it demands is rational.*

But while it is easy to see that this retort is tempting,[14] it won't do. For (**RISC**'') governs what it is rational for each person to believe, not what is transcendently true. The Enlight-

13. If there is any tendency to doubt this, consider the Cartesian assumptions listed above.
14. I read in my school days a passage in Malebranche, which I haven't been able to find again, but of which I remember the content as going roughly thus:

enment conception of *universal reason* cannot dictate that any given argument must secure a single reaction in all audiences, regardless of their epistemic background. What may be salvaged from it is a conception of the role of reason as *negative*: as able, like Socrates's eristic method, not to reveal truth but to expose falsehood, or at least *rational unacceptability to a person at some time*. This negative view of reason is not the *minimal* view: some at least, notably Hume, have claimed even less for reason: have claimed, indeed, that reason can't be relied on even to eliminate falsehood. In that light, the negative view is far from pessimistic. Negative reason remains a modest but not insignificant legacy of the Enlightenment, which philosophical stupidity can endorse in good conscience.

Coda: Amends to Descartes

I end with a note intended to correct the impression I may have given, that I regard Descartes as a villain in the struggle between oppressive cleverness and subversive stupidity. Actually Descartes, like Socrates, is among the heroes of philosophical stupidity. For while he professes all sorts of principles that seem to me absurd, he is also prescribing that my own *assent* to those principles is what must ground my further convictions. To be sure, he expects that by following his meditation in my own mind, I shall arrive at the same conclusions; further, he expects this because these opinions are dictated by the Light of Nature. But *from my own subjective point of view*, the fact that these opinions are dictated by the Light of Nature is perceptible only through the prism of their appearing indubitable *to me*. No separate process I can go through will distinguish what is true from what merely appears true. So if I take care to follow the meditative course Descartes prescribes, I will be doing just what is required by the ideal of philosophical stupidity I have been recommending. For despite all his humble protestations of submission to the Church, Descartes implies that, under the guidance and discipline of my individual consciousness, my own subjective judgment is the ultimate measure of God himself.

References

Bencivenga, E. 1993. *Logic and other nonsense: The case of Anselm and his God*. Princeton: Princeton University Press.

Burkert, W. 1996. *Creation of the sacred: Tracks of biology in early religions*. Cambridge, MA: Harvard University Press.

Cohen, J. L. 1981. Can human irrationality be demonstrated? *Behavioral and Brain Sciences* 4.

Some people, who had erroneous ideas and then gave them up, expect those whose beliefs are correct to give them up just as easily. But what they fail to see is the vast difference between firmness of mind, which is the virtue of those who are steadfast in true beliefs, and the mere obstinacy of those who cling to false doctrines.

If nothing like this is in Malebranche, then let my false memory bear witness to my own occasional temptation to draw such a "distinction". Others may conceivably have felt it too.

Davidson, D. 1982. *Inquiries into truth and interpretation*. Oxford: Oxford University Press, Clarendon.

Dawkins, R. 1982. *The extended phenotype: The gene as unit of selection*. Oxford : Oxford University Press.

Descartes, R. 1964–76. *Oeuvres de Descartes*. Textes présentés par C. Adam and P. Tannery. Paris: Vrin/CNRS.

Gardner, M. 1957. *Fads and fallacies in the name of science*. New York: Dover.

Gewirth, A. 1998. *Self-fulfillment*. Princeton: Princeton University Press.

Gigerenzer, G., P. Todd, and ABC Research Group. 1999. *Simple heuristics that make us smart*. New York: Oxford University Press.

Glover, J. 1999. *Humanity: A moral history of the Twentieth Century*. London: Jonathan Cape.

Gosse, P. H. 1857. *Omphalos: An attempt to untie the geological knot. London: J. van Voorst.*

Hintikka, J. 1962. "Cogito, ergo sum": Inference or performance? *Philosophical Review* 71: 3–32.

Hume, D. 1975. *Enquiry concerning human understanding*. 3d ed. Ed & introd by L. A. Selby-Bigge. Revised by & notes by P. H. Nidditch. Oxford: Oxford University Press, Clarendon.

Kahneman, D., P. Slovic, and A. Tversky, eds. 1982. *Judgment under uncertainty: Heuristics and biases*. Cambridge and New York: Cambridge University Press.

MacIntyre, A. 1981. *After virtue: A study in moral theory*. Notre Dame, Indiana: University of Notre Dame Press.

Matthen, M. Forthcoming. Two ways of thinking about fitness and natural selection.

Nozick, R. 1981. *Philosophical explanations*. Cambridge, Massachusetts: Harvard University Press, Belknap.

Quine, W. V. O. 1960. *Word and object*. Cambridge, Massachusetts: MIT Press.

Sen, A. 2000. East and West: The reach of reason. *The New York Review of Books* 47(12), 20 July.

Skinner, B. F. 1948. 'Superstition' in the pigeon. *Journal of Experimental Psychology* 38: 168–72.

Sober, E., and D. S. Wilson. 1998. *Unto others: The evolution and psychology of unselfish behavior*. Cambridge, MA: Harvard University Press.

Taylor, C. 1989. *Sources of the Self*. Cambridge, MA: Harvard University Press.

Toulmin, S., and J. Goodfield. 1990. *Cosmopolis: The hidden agenda of modernity*. New York: Free Press.

van Inwagen, P., and D. Hill. 2000. Peter van Inwagen interviewed by Daniel Hill. *Philosophy Now* 24: 27–29.

Vlastos, G. 1966. Zeno's race course. *Journal of the History of Philosophy* 4:95–108.

Wittgenstein, Ludwig. 1969. *On Certainty*. Oxford: Basil Blackwell.

Rationality and Irrationality in Scientific Language

LUIS FLORES H.*

1. Introduction

Let us define "rationality" as the property (of statements, actions, facts) of being justified (by other statements, actions, facts). This justification can be epistemic (arguments), deontic (norms), or it can take some other form. Science concerns itself with ends like verisimilitude and with means like methods and procedures. If the appropriate relation between ends and means obtains, then statements, actions and facts are justified in science. In this sense, science is rational. Scientific rationality can find its expression in different sorts of scientific theories and languages. Natural sciences employ a specialized natural language (English, German, French, and so on) as well as formal symbolic systems (geometry, algebra, etc.) and other semiotic systems such as chemical notation. Since scientific language comprehends a broad spectrum, I restrict myself here to the language of the natural sciences.

My first claim is that there are at least four kinds of scientific rationality: the *cognitive*, in virtue of which science is able to describe the world; the *symptomatic*, in virtue of which we are able to grasp the mental states of scientific speakers; the *deontic*, in virtue of which we are able to change the behavior of persons and animals; and the *aesthetic*, in virtue of which science is evaluated according to the idea of beauty. Consequently, scientific language, in addition to having a cognitive function, is marked also by symptomatic, deontic, and aesthetic functions. But more: these functions constitute a modifiable network because they interact holistically in a synchronic and diachronic way. In this sense, the cognitive and deontic functions of scientific language may be said to prevail in normal science, whereas the symptomatic and aesthetic functions take on a dominant position in the turning points of scientific change. This paper thus criticizes the common philosophical view which sees the language and hence the scientific rationality of natural science entirely in terms of the cognitive function. On the one hand, this view – which we might call a *cognitive approach* – is dominant in the philosophy of science and conceives scientific languages either from the point of view of logic (for example as regards logical consistency) and mathematics (in terms of set-theoretical formalism) or from the point of view of history and sociology. On the other hand, the philosophy of language has not been especially concerned with scientific languages.

The cognitive approach was initiated by Carnap: "Science is the system of *intersubjectively valid statements*"(Carnap 1995, 96). Popper also defends this approach, holding that in language there are four functions: the expressive, the signaling, the descriptive, and the argumentative functions. But, as regards scientific language, he considers the first two of these as lower-level functions (remarking that there are "some other functions that play no role here, such as a hortative and a persuasive function" (Popper 1972, 235)). In our view, all four functions are of equal importance, the descriptive and argumentative functions being subfunctions of the cognitive function. Furthermore, the hortative and the persuasive functions play a role in the context of scientific discovery.

Kuhn introduces the concept of argumentative persuasion in scientific language, but for him science is always a matter of "scientific knowledge" (Kuhn 1972, 210). Van Fraassen has extended this approach with his pragmatics of explanation and with his attempt to bring together the model structures found in semantics with the models of scientific theories (van Fraassen 1980). From the viewpoint of the cognitive study of science, R.Giere holds that: "The more promising course for the cognitive study of science would seem to be investigating the roles played by the specific cognitive mechanisms of representation and judgment in scientific research" (Giere 1987, 153). Recent discussions consider whether there is a reference for scientific terms (Mülhölzer 1995). In all these cases, we are still in the field of the cognitive function. The problem with all of these approaches, then, is that aspects of science and scientific processes that fall outside the scope of cognition are conceived, in effect, as irrational.

My second claim is that irrationality has a role in natural science, and thus in the language of natural science, in the sense that there is a certain sort of relative irrationality which can arise within science in opposition to a specific type of scientific rationality. What Newton-Smith calls the "temperate rationalist" has difficulty accepting relative irrationality: "But, unlike the rationalist, he is interested in explaining in terms of non-scientific interests why the rational pursuit of science should have become a human interest " (Newton-Smith 1981, 273). Newton-Smith's view represents what we might call a *rationalist approach.* From this standpoint scientific irrationality can be complementary to scientific rationality, and it takes place especially in phases when science endures one or other type of developmental crisis.

2. Cognitive Rationality at Work in Scientific Language

Let us consider the cognitive function of scientific language. The rational principles concerning scientific language imply that there must be an agreement among the users of this language both as regards the use of symbols and also as regards the privilege of the cognitive function over the other functions. All semiotic systems of scientific languages are oriented toward the value of truth. It is for this reason that the cognitive function is dominant in scientific language, while the other functions are recessive.

The *cognitive* function calls for three types of statements in the natural sciences: descriptive statements involving meaningful signs that denote empirical objects, operational statements involving signs whose function is to indicate measuring operations, and constructive statements involving meaningful signs that denote abstract objects. This is the cognitive form of scientific rationality. The acceptance of such statements is either *rational* or it is not, depending on arguments, experimental evidence, reasons, and so on.

* Professor of the P. Catholic University of Chile. This paper was written as part of the FONDECYCT Research Project N°1990520. I am grateful to Prof. Barry Smith and to my referees for their criticism and suggestions.

B. Brogaard, B. Smith (Hg.), *Rationality and Irrationality / Rationalität und Irrationalität*, S. 94–100.

3. Some Consequences of Hard and Soft Cognitive Rationality in Scientific Language

There is a progression which we can witness in history from figurative (iconic) language toward characteristic (arbitrary) language (Granger 1993, 54). For example, alchemy often used images to depict natural elements. As alchemy gave way to chemistry, these icons were replaced by arbitrary (non-figurative) signs. This is the triumph of what Leibniz called the *characteristica universalis.*

Polysemy characterizes natural languages, but to the degree that natural language has become specialized in the natural sciences such polysemy is frustrated through the requirement that each symbol has one and only one meaning, a requirement which becomes an ideal in scientific language. From the historical viewpoint, this ideal, too, is related to Leibniz and his *characteristica universalis.* Later, Frege defends it, maintaining that the sense and reference of terms should be invariable in a perfect language (Frege 1892, 198). This is part of Frege's ideal of an ideography *(Begriffsschrift).*

Deixis also tends to disappear in scientific language because there is a tendency to eliminate any allusion to the context of use. In a statement of Pythagoras' theorem or of Newton's theory of gravity, it does not matter who speaks, or when, or where, or to whom. Cognitive rationality requires abstraction from the context of enunciation.

It is true that the concepts used in science represent the progressive elimination of metaphor. But, like the phoenix, metaphor is also always reborn. The metaphor of air as a sea fulfilled a heuristic function because it allowed Torricelli to pass from hydrology to pneumatics. Harvey did something similar when he described the heart as a bellows. Mark Johnson points to: "the growing body of literature on the pervasiveness and indispensability of metaphor in science ... Many have become convinced that there can no longer be facile dismissal of the cognitive importance of metaphor." (Johnson 1981, 42). Relative to earlier reductionistic conceptions of metaphor entirely in terms of their aesthetic function, the recognition of the cognitive function of metaphor was an advance. This still does not go far enough, however, because metaphor has also a deontic and a symptomatic function.

There is an opposition between language *(langue)* and speech *(parole)* in scientific language. The former has a privilege over the latter because of the requirement placed on scientific language that it be universal. The passage from Euclidean geometry to non-Euclidean geometry and then to metageometry attests to this search for universality; it is cognitive rationality at work. But even so, the particular cannot always be banished. In quantum mechanics, Heisenberg uses infinite matrices as notation, whereas Schrödinger uses the wave equation as notation. We see here an essential tension between universality and individuality.

The opposition between synchrony and diachrony transforms itself into the negation of the latter. This explains, to a certain extent, the appeal of dead languages (Greek, Latin), particularly in taxonomies. The terminology should not change, but this ideal applies only within the framework of Kuhn's normal science, and we need to consider scientific revolutions and the changes of scientific meanings which these involve.

Symbolic economy is taken to the extreme in science. Such economy is of course manifested already in natural language, but abbreviations (ACTH, for adrenocorticotrophic hormone), acronyms ("laser"), and so on, are systematized in scientific language. This too represents a victory of what Leibniz called blind or symbolic thought, especially in the case of mathematics. Here cognitive rationality means the economy of symbolic means.

Scientific language seeks accuracy. Invention of technical words secures this quest: "bug" becomes "fly" and later *"musca domestica"* in entomology. Here cognitive rationality means the making more precise of symbols.

All these properties belong to what we might call a *hard cognitive rationality* in the language of natural science, that is to say, to cognitive rationality as an ideal. But real science is not like this. For instance, nowadays vagueness in scientific language does not astonish us. As MacCormac remarks: "Newton's ambiguous and metaphorical use of the term 'force' does catch us unawares. For those who have assumed that Newtonian mechanics was a paradigm of logical rigor and a model for how scientists should construct theories, it is absolutely shocking!" (MacCormac 1976, 35). But even the recognition of metaphor and vagueness is not enough; it is not only that there is what we might call a *soft cognitive rationality* in the language of natural science. There is something more.

4. Other Types of Rationality at Work in Scientific Language

The *symptomatic* function tends to be eliminated from scientific language, which generally refrains from indicating personal experiences. Scientific prose eliminates the diminutive, the insult, the interjection. That is why the symptomatic function is confined to a scientist's journal or correspondence. Nevertheless, we do find it in scientific texts, when it is necessary not only to convince, but also to persuade. In this case, we take into account the states of mind of different types of interlocutors. The aim here is eloquence of scientific language, and the means are expressive. This is symptomatic rationality, and includes the rhetorical dimension of science. Again, we would be wrong to reduce such rhetoric to its merely cognitive role, as does Gross, who is "committed to the view that rhetoric has a crucial epistemic role in science, that science is constituted through interactions that are esssentially rhetorical " (Gross 1990, x). There is a sort of scientific use of the patronymic, a special case of the symptomatic function, which appears for example in statements like: "Malpiguian corpuscle", "Malpiguian layer" and "Malpiguian tuft" and so forth. These refer not merely to the corresponding anatomic object but also to Marcello Malpigui, the Italian physician and anatomist of the seventeenth century who was their discoverer. In the pancreas there are cells called "islets of Langerhans". This statement combines a geographical metaphor having a heuristic function with a historical connotation related to Paul Langerhans' act of discovery. Newton's theoretical terms "absolute space" and "absolute time" are not only concepts required for understanding relative space and relative time; they have also what I would like to call a theophanic connotation of being the *sensorium Dei* (another case of the symptomatic function at work). These historical or cultural connotations belong to the context of discovery, but not to the context of justification. This notwithstanding, however, they should be included in the analysis of scientific language.

The *deontic* function is restricted to directions for real or mental experiments

(*Gedankenexperimente*). To follow these directions is either *reasonable* (in contrast to *rational*) or it is not. The imperative mode of scientific language is significant here. This is deontic rationality at work aiming, in the natural sciences, at the limitation of the possible range of new events , whether real or mental. Bunge recognizes that we find *proposals, problems* and *rules* in scientific languages and that these are neither true or false. Moreover, he says that we do not find in the body of scientific language "*advices* ..., *requests* ... and *commands*." He points out that we do meet "these kinds of objects ... in the course of research, as of any other action, but not in the outcome of research." (Bunge 1967, 51–52). Unfortunately however this reduction to the context of justification is an obstacle in the way of a better understanding of the language of natural science. We should take account also of the context of scientific discovery.

The *aesthetic* function is confined to the architecture of symbols (e.g., the beauty of Maxwell's equations). Kuhn recognizes the presence of this function in science: "the importance of aesthetic considerations can sometimes be decisive." (Kuhn 1970, 156). He does not, however, develop this view systematically. The criteria for evaluating the aesthetic qualities of scientific language are simplicity, order, harmony, symmetry, unity, isomorphism, elegance, and so on. Aesthetic rationality is ruled by these criteria. Accordingly, scientific language is marked by the feature of *ratio* (proportion, measure). When A. I. Miller examines some aesthetic aspects of Poincaré's and Einstein's thinking, he does not necessarily suppose that these play a cognitive role: "For example, can aesthetic sensibility be defined operationally post hoc as a heuristic or should it be considered a mental algorithm?" (Miller 1984, 240). Unfortunately however he does not examine the influence of this aesthetic sensibility in modelling scientific language.

A general argument for my thesis concerning this network of four functions in scientific language is that the latter can be conceived as both product *(poiesis)* and action *(praxis)*. On the one hand, as *product* there is a network of true or false statements, and this can also serve aesthetic criteria (it can have a certain style). On the other hand, as *action* there is a network of theoretical and experimental practices that, in accordance with the deontic function, depend upon the freedom of scientists and their responsibility to be truthful. This action can also be, in accordance with the symptomatic function, an indication of the mental states of the scientists involved, for example when a question indicates a vacillating state of mind. Furthermore, processes concerning aesthetic, deontic and symptomatic functions in scientific language are not merely chaotic, because they can be justified and, accordingly, be rational. On the one hand, natural science has to do with natural reality and this has not only cognitive, but also aesthetic aspects. On the other hand, natural scientists are human beings and, therefore, there are deontic and symptomatic aspects in their search for natural reality.

However, these kinds of rationality are not mutually exclusive but rather complementary, according to the synchronic level or the diachronic phase considered. During phases of normal science in Kuhn's sense, cognitive and deontic rationality are prevalent: they provide orientation for scientists in the refinement of their paradigm and in the elaboration of appropriate experiments . At the moment of crisis, however, when one paradigm begins to give way to a new one, symptomatic and aesthetic rationality are dominant. Why? Because the shift of a paradigm presupposes the Kuhnian "conversion" of the scientists making up a particular scientific community, because the rhetoric of science has to

do with speakers and audiences, and because the new world view, so long as it remains, as knowledge, still uncertain, needs formal criteria more concerned with the aesthetic function. Thus that shift is more involved with symptomatic and aesthetic rationality. These changes in scientific setting are also essential to my proposal.

Consider the Newtonian metaphor of the universe as clock. This has a symptomatic dimension because we are able to grasp the background metaphor of the universe as mechanism. It has a heuristic dimension in that it yields a model for depicting physical reality, and it has also an aesthetic dimension relating to the simplicity of a clock. Finally, the deontic function of this metaphor is to provoke a special way of thinking about physical reality.

4. Rationality and Irrationality in Scientific Language

During periods of what Kuhn calls paradigmatic crisis, different schools or trends appear. They speak different dialects which are marked by an incommensurability of meanings. Scientists' machinations go beyond the four kinds of rationality mentioned above. Scientific language becomes a Tower of Babel, an example of irrationality. In this phase, polysemy, particularities, redundancy, changeability, and inaccuracy become even stronger in the symbolic domain of natural science than any soft cognitive rationality. Moreover, some terms may arise which are extrapolated from other disciplines (for instance, "genome" employed in relation to a linguistic code) and are used without sufficient knowledge, or are laden with unscientific content (Galileo, for example, perhaps out of Neoplatonic motives, considered the circle to be perfect and saw Kepler's ellipse as something degenerate). But this irrationality is sometimes necessary if a theoretical leap is to be possible. In this sense, it is not an *absolute* irrationality, but rather something that has value only *relative* to a certain kind of overarching rationality. Statements, facts and actions that exist out of the accepted scientific paradigm are irrational because they cannot be justified by this. Nevertheless, some of them can be the progenitors of a new theory or even of a new paradigm. The rationality and irrationality of the language of natural science are systole and diastole in scientific flux.

5. Conclusions

The orientation toward hard cognitive rationality was decisive in binding the language of natural sciences to the ideals of universality, arbitrariness, unambiguousness, abstraction, fixity, economy, accuracy and truth. But these are more ideals than reality. Today we accept a softer cognitive rationality in the language of natural science.

We should extend our cognitive understanding of the rationality of scientific language to include symptomatic, deontic and aesthetic rationality in a holistic network.

And finally we should understand the rationality of scientific language as being complementary to a certain sort of irrationality. This is because, when we consider the history of natural sciences and the rise and fall of scientific theories, we need to recognize that this is a sequence in which periods of Kuhnian normal science alternate with periods of

crisis and of paradigm shift.

Briefly, my thesis puts forward a new way of understanding the complexity of the language of natural science, for that reason it excludes both cognitive and rationalist approaches.

Literature

Bunge, M. 1967 *Scientific Research I. The Search for System,* Berlin: Springer.

Carnap, R. 1995 *The Unity of Science,* Bristol: Thoemmes Press.

Frege, G. 1892 "On Sense and Nominatum", in A. P. Martinich (Editor), *The Philosophy of Language*, New York Oxford: Oxford University Press, 186–198.

Giere, R. N. 1987 "The Cognitive Study of Science", in N. J. Nerssesian (Editor), *The Process of Science,* Dordrecht: Martinus Nijhoff Publishers.

Granger, G. G. 1993 *La science et les sciences,* Paris: Presses Universitaires de France.

Gross, A. G. 1990 *The Rhetoric of Science,* Cambridge: Cambridge University Press.

Johnson, M. 1981 *Philosophical Perspectives on Metaphor*, Minneapolis: University of Minnesota Press.

Kuhn, T. S. 1972 *The Structure of Scientific Revolutions*, Chicago: The University of Chicago Press.

Miller, A. I. 1984 *Imagery in Scientific Thought,* Boston: Birkhäuser.

Mühlhölzer, F. 1995 " Science without reference? ", *Erkenntnis,* 42, 203–222.

MacCormac, E. 1976 *Metaphor and Myth in Science and Religion,* Durham: Duke University Press.

Newton-Smith, W. H. *The Rationality of Science,* Boston: Routledge & Kegan Paul.

Popper, K. 1972 *Objective Knowledge,* Oxford: Oxford University Press.

van Fraassen, B. C. 1980 *The Scientific Image,* Oxford: Oxford University Press.

Oxford Philosophy: A Case Study in Cognitive Epidemiology

LYND FORGUSON

During the quarter century following the end of World War II, Oxford University placed more of its philosophy graduates in teaching positions, both in Great Britain and elsewhere, than any other university, before or since. The period of Oxford's dominance of the philosophy job market coincided almost exactly with the heyday of a distinctive philosophical movement, most familiarly known as "ordinary language philosophy," but also known as "linguistic philosophy," "linguistic analysis," "conceptual analysis," or simply "Oxford philosophy," acknowledging the fact that most of those who philosophized in this style in the post-war period were either philosophy teachers in Oxford University or had received at least part of their formal philosophical education there.

Philosophical movements share many of the characteristics of epidemic diseases, which typically arise among a few individuals in a particular location, and soon spread throughout a wider population. The rapid spread of a disease causes alarm; attempts are made to eradicate it, or at least to control the contagion. After a period of widespread infection, the disease generally subsides.

The discipline of epidemiology investigates the incidence, distribution and control of diseases in a population. Among the leading questions of epidemiology are these. Where did the disease originate? By what means is it transmitted? What accounts for its particular geographical distribution, and the rate at which it spreads through the population? Very similar questions arise when a cultural phenomenon such as a philosophical movement is the focus of enquiry. How can we best account for its geographical incidence and distribution, and for the rate at which it spread beyond its location of first incidence? These are broadly empirical questions, and I think a broadly epidemiological methodology is appropriate for pursuing answers.

The information available for an epidemiological investigation of ordinary language philosophy includes, of course, the standard sources of information about a philosophical movement: the philosophical publications of the ordinary language philosophers and their opponents, including both their philosophical writings and also any biographies, autobiographies or memoirs. But in addition to these, and particularly likely to be useful, are institutional records, such as archives of examination questions and lists of teaching staff and where they were educated. These can reveal interesting correlations between the career movement of philosophers and the spread of ordinary language philosophy.

Ordinary language philosophy grew out of Ludwig Wittgenstein's recognition, in about 1928, that the "picture" theory of meaning he had set forth in his *Tractatus* was inadequate. Upon his return to Cambridge later that year, he began to work out a radical rethinking of the role of language in philosophical enquiry, one that emphasized the way that philosophical problems arise "when language goes on holiday." Our inattention to the way words are actually used in the various language games in which we engage gives

B. Brogaard, B. Smith (Hg.), *Rationality and Irrationality / Rationalität und Irrationalität*, S. 101–109.

rise to philosophical perplexity, which in turn leads to the construction of philosophical theories. Once we come to understand the way our words are ordinarily used, the philosophical problems dissolve, and the urge to philosophize subsides.

Wittgenstein died in 1951, without publishing anything of his later philosophy, though some of its flavour was conveyed in publications during the later 1930s by his followers John Wisdom and G.A. Paul, and also through clandestine circulation of the "Blue" and "Brown" books. By the time his *Philosophical Investigations* came out posthumously in 1953, a similar philosophical approach was already well-established at Oxford, and had reached a wide audience through the publication of a number of articles by J.L. Austin and a group of younger Oxford philosophers whom he had influenced, and also through the publication in 1949 of Gilbert Ryle's *The Concept of Mind*. The label "ordinary language philosophy" came to be applied indiscriminately to the work of the later Wittgenstein and his followers, as well as the work of the Oxford group and their sympathizers, as though they formed a single philosophical approach. Indeed, their similarities are more salient than their differences.

The Oxford ordinary language philosophers shared with Wittgenstein the view that many, if not all, of the traditional philosophical problems, as well as the theories that have been put forward in response to them, have their genesis in an inadequate grasp of what Ryle called "the logical geography" of the linguistic expressions we employ in our everyday understanding of ourselves and the world around us. Austin gave it a Darwinian spin:

> Our common stock of words embodies all the distinctions men have found worth drawing, and the connexions they have found worth marking, in the lifetimes of many generations: these surely are likely to be more numerous, more sound, since they have stood up to the long test of survival of the fittest, and more subtle, at least in all ordinary and reasonably practical matters, than any you or I are likely to think up in our arm-chairs of an afternoon – the most favoured alternative method (Austin 1956).

The first task of philosophical enquiry, therefore, should be to describe as carefully and as patiently as possible the way we actually do or would in ordinary life use the linguistic expressions germane to areas of traditional philosophical contention.

Unlike Wittgenstein and his followers, though, Austin and Ryle and the philosophers close to them did not think that the sole role of philosophy is the therapeutic removal of philosophical perplexity. Linguistic philosophy not only has the critical task of clearing away the debris of confused theorizing; it also has the constructive task of clarifying our concepts and their interrelationships.

Though the Oxford and Cambridge strains of ordinary language philosophy are distinguishable, there is epidemiological evidence linking Wittgenstein's thought to the development of the Oxford strain. Ryle first met Wittgenstein at a philosophical conference in 1929. They formed a friendship and occasionally went on walking holidays together throughout the 1930s, (Monk 1990) during which times they are likely to have had philosophical discussions. Until he met Wittgenstein, Ryle had been attracted to Phenomenology, but in 1932 he published "Systematically Misleading Expressions," a paper thoroughly linguistic in approach, foreshadowing some of the themes of *The Concept of Mind*, as do several of his other papers of the 1930s. Many commentators have also remarked on the close similarity between the "logical behaviourism" of *The Concept of Mind* and *Philosophical Investigations*.

Austin's brand of ordinary language philosophy was formed in the later 1930s, during regular discussions in Oxford with Isaiah Berlin and A.J. Ayer, later expanded to include Stuart Hampshire and a few others. According to Berlin, in a memoir written nearly forty years later, Austin already at that time

> believed that the only reliable method of learning about types of action, knowledge, belief, experience, consisted in the patient accumulation of data about actual usage. Usage was certainly not regarded by him as sacrosanct, in the sense of reflecting reality in some infallible fashion, or of being a guaranteed nostrum against confusions and fallacies. But it was neglected at our peril (Berlin 1973).

Ayer was already famous as the author of the recent bombshell, *Language, Truth and Logic* (Ayer 1936), and the meetings soon developed into a running debate between Ayer and Austin, the former championing the logical positivism he had acquired attending meetings of the Vienna Circle in 1932-33, the latter trenchantly attacking what he took to be the crude oversimplifications of the positivist critique of philosophy.

> These discussions led to the emergence of 'Oxford Analysis,' not so much as a consequence of Austin's specific theses, as from the appeal to common linguistic usage which was made by us all, without, so far as I recollect, any conscious reference at the time to Wittgenstein's later doctrines. ... Certainly [Austin's] first published contribution to philosophy – the paper on 'A Priori Concepts' [Austin 1939] in which a good deal of his positive doctrine is embodied – owes ... nothing to any acquaintance with Wittgenstein's views, unless perhaps, very indirectly, via John Wisdom's articles which he certainly read. (Berlin 1973).

Despite Berlin's disclaimer, however, there is evidence that Ayer had already been exposed to the basic thrust of Wittgenstein's new approach during his brief stay in Vienna, and he in turn had conveyed this to his friend Isaiah Berlin in a letter from Vienna in 1933.

> Wittgenstein is a deity to them all, not mainly on the strength of the *Tractatus*, which they consider a slightly metaphysical work . . . but on the ground of his later views. . . Philosophy is grammar. Where you would talk about laws they talk about rules of grammar. All philosophical questions are purely linguistical. And all linguistical questions are resolved by considering how the symbol under consideration is in fact *used* (Rogers 1999).

After the war, Wittgenstein's influence could also be felt in Oxford through the presence of Friedrich Waismann, Elizabeth Anscombe and G.A. Paul, and bootlegged copies of the "Blue" and "Brown" books were by then in wide circulation. There were, then, both during the 1930s and in the post-war years prior to the publication of *Philosophical Investigations*, ample opportunities for Oxford-based philosophers and students to gain some

exposure to Wittgenstein's later philosophy. Of the two most influential Oxford philosophers of the period, Ryle probably absorbed more of Wittgenstein's philosophy than he ever himself realized. Austin's philosophy is quite unlike Wittgenstein's, apart from a shared conviction that ordinary language is our chief resource for philosophical analysis. The main reason that Oxford philosophy developed in its own distinctive way, in some respects antithetical to Wittgenstein's vision, was that the chief *animateur* of the movement was always Austin, and not Ryle, as important as Ryle's contribution to philosophy was.

At any rate, it was Austin and Ryle's Oxford, and not Wittgenstein's Cambridge, that was chiefly responsible for the rapid spread of the ordinary language movement after the war. When the war ended, the relatively small British university system strained to absorb a multiple cohort of students. Those just reaching university age in the immediate post-war years were joined by returning veterans whose education had been interrupted by the war effort. At Oxford the student population grew by over 60% in the two year period between 1946 and 1948. The initial enrolment bulge tapered off in Britain by 1950, but the proportion of school-leavers going on to higher education increased gradually during the 1950s, and by the early 1960s, the "baby boom" generation began to reach university age, causing a second rapid bulge in the system. The increase in student numbers naturally led to a corresponding increase in demand for university teachers in all fields, including philosophy. Because Oxford annually produced far more philosophically-trained graduates than any other British university, Oxford was in a position to dominate the job market throughout the long period of expansion.[1]

Year	1939	1950	1955	1960	1965	1970	1975
British Universities	25	26	26	28	32	42	45
Oxford Philosophers	11	43	46	48	52	59	60
Other UK Philosophers	78	127	135	176	288	363	416
Total UK Philosophers	89	170	181	224	340	422	476
Philosophers with Oxford degrees in the rest of UK	28	68	80	90	133	165	173
of which B.Phil.	0	6	13	22	68	103	121

Table 1. traces the growth of the British university system from 1939 to 1975, and the corresponding growth in the number of professional philosophers. It also indicates the growth of the philosophy contingent within Oxford, and the impact of Oxford-trained philosophers on the job market.[2] By 1950, there were almost twice as many philosophers

1. About 20% of all undergraduates at Oxford follow a program of studies involving a concentration in philosophy. In these early post-war years, a First Class B.A. was sufficient to secure a teaching position in a British university, and Oxford itself made most of its junior appointments from among its own recent graduates.

teaching in British universities than there were on the eve of the war. The profession grew by another 32% during the next ten years, and by another 50% by 1965. In the quarter century between 1950 and 1975, the number of professional philosophers in Britain had increased nearly threefold. Oxford graduates accounted for 44% of the profession in 1939. If one subtracts from the total those holding positions at Oxford, over 30% of philosophers in the rest of the country's universities were Oxford graduates. The proportion of Oxford graduates teaching philosophy in other British universities reached a peak of 59% in 1955. But even twenty years later, at a time when many other British universities were competing with Oxford for a share of the job market, the graduates of this one university made up over 40% of philosophers teaching in universities other than Oxford. If one adds in the philosophers on Oxford's own teaching staff, the proportion of Oxford graduates rises to nearly 50% of the total.

Oxford's own teaching strength in philosophy had to be increased rapidly in the early post-war years to accommodate the sudden influx of students. Ryle and Austin were among the pre-war group of ordinary language philosophers who now returned to resume interrupted careers. Ryle was appointed Waynflete Professor of Metaphysics in 1945, and Austin became White's Professor of Moral Philosophy in 1950; thus two of the three university professorships in philosophy were held by the leaders of the ordinary language movement. Many of the new appointments made during these early post-war years also went to young philosophers sympathetic to this approach, so that by 1950 there was a "critical mass" of about fifteen ordinary language philosophers active in Oxford, and these in turn were in a position to influence the teaching of philosophy in the university throughout the 1950s and beyond.[3]

There was, in these early postwar years, an aura of revolutionary excitement still recalled by many of those who studied at Oxford during this time.

> One had something of the sense of a new beginning being made in the field, of breaking totally fresh ground, of an intellectual renaissance, akin to that of Descartes or the introductory pages of Hume's writings. After centuries of muddle and blunder and confusion and pedantry and obfuscation, at last a new age had dawned. 'Bliss was it in that dawn to be alive.'[4]

One indicator of the growing influence of the ordinary language approach on the

2. All information in Table 1 is from the *Yearbook of the Universities of the British Empire* (1914-). The number of philosophers listed for Oxford in 1939 includes only the three professors and those holding university lectureships. However, the *Oxford University Calendar* for that year lists 37 members of the Sub-Faculty of Philosophy, including those holding appointments in Oxford's teaching colleges, as well as those with university-wide appointments.
3. Gilbert Ryle, J.L. Austin, G.J. Warnock, J.O. Urmson, H.P. Grice, P.F. Strawson, Stuart Hampshire, D.F. Pears, R.M. Hare, P.H. Nowell-Smith, T.D. Weldon, Isaiah Berlin, Friedrich Waismann, G.A. Paul, and G.E.M. Anscombe differed from one another in many ways, but were generally identified as ordinary language philosophers by others if not by themselves. Ryle, Austin, Berlin, Weldon, Hampshire, and Grice had been members of the philosophical community of Oxford before the war. The others were appointed in the five years between 1946 and 1950.
4. David Gallop, letter to the author, 1998.

philosophical climate in Oxford during these years is the increasing frequency with which questions inviting an ordinary language approach appeared on the undergraduate final examination papers in philosophical subjects. As early as 1947, the following question appeared: "Is it important for the philosopher to attend to the everyday use of language?" Ten years later, eleven of fourteen questions on one of the examination papers reflected in one way or another the current concerns of the ordinary language philosophers. I do not mean to suggest that candidates were at a disadvantage in the examinations if they did not adopt an ordinary language approach in their answers, but to do well they would at least have had to grapple with them in a way that would reflect their familiarity with the issues raised by the questions, and with the publications to which their tutors had directed them.

Oxford introduced a new graduate degree after the war, the Bachelor of Philosophy, which required a series of examinations and the preparation of a short thesis. Specifically designed to prepare its graduates for teaching careers in philosophy, the first B.Phil. degrees were awarded in 1948, and by 1965, almost a quarter of all philosophers teaching in British universities other than Oxford were B.Phil. graduates. (See Table 1.) Questions reflecting the interests of ordinary language philosophers were as prominently featured in the B.Phil. examinations as they were in the B.A. examinations, from the first set of examinations in 1948 until the early 1960s.[5]

One measure of the effect ordinary language philosophy had on the conduct of philosophical enquiry in Britain during this period is provided by a survey of the publications of the B.Phil. graduates. Publications were increasingly becoming important for career advancement during this period, so if ordinary language philosophy was indeed "carried" throughout Britain by Oxford-trained philosophers, one would expect this to be reflected in their publications.

If one concentrates on those who were employed in British universities at any time between 1950 and 1965, one finds that no more than 40 of them ever published anything that could reasonably be identified as exhibiting an ordinary language approach, and 10 of these held appointments at Oxford.[6] The rest were distributed among 20 institutions. There were, of course, ordinary language philosophers among those who were not B.Phil.s, but the total number of ordinary language philosophers in British universities other than Oxford during this period never exceeded about 20% of the total.

From an epidemiological perspective, though, these are impressive figures. In 1965 there were 32 institutions in Great Britain offering instruction in philosophy. Nearly two-thirds of them harboured at least one carrier of the ordinary language disease; and of

5. B.A. and B.Phil. examination papers are on deposit in the Bodleian Library, Oxford University.
6. The *Oxford University Calendar* listed B.Phil. degrees awarded each year from 1948 through 1955. Since 1956, the names of those awarded this degree each year have been listed in the *Oxford University Gazette*. Information on the publications of Oxford B.Phil. graduates is taken from *The Philosopher's Index* (1940-). In determining whether a listed publication should be classified as reflecting a broadly "ordinary language philosophy" orientation, I looked at the abstract published with the *Philosopher's Index* entry. If no abstract was included, I read the listed publication if possible. In some cases, I made a judgement based on the publication's title. If the title did not make it obvious that the author based philosophical conclusions on an appeal to ordinary language I did not count it.

the total population of 340 philosophers teaching in Britain in that year, at least ten, and perhaps as many as twenty per cent of them appear to have been infected with ordinary language philosophy at some time during the fifteen year period. If we were faced with a real disease, and not merely metaphorically characterizing a philosophical movement in those terms, we would have ample grounds to think that we had an epidemic to contend with.

Ordinary language philosophy was not confined to Great Britain. It made inroads in other English-speaking countries, including the United States and Canada, though nowhere outside Great Britain was its presence more widespread, at least for a brief period, than in Australia. During the period covered by this investigation, generous government scholarships were available for Australian students to pursue graduate studies abroad. In philosophy, a substantial number of them came to Oxford to read for the B.Phil., and most of them seem to have returned to Australia, where the university system was growing as it was in Britain and elsewhere. Melbourne University had become an outpost of the Wittgensteinian variety of ordinary language philosophy as early as the 1930s, with the arrival of George Paul, and later Douglas Gasking, both of whom had studied with Wittgenstein at Cambridge. As late as the mid-1960s, final examination papers at Melbourne were dominated by questions requiring the student to interpret or discuss specific passages from *Philosophical Investigations* or other of Wittgenstein's published works, but by 1955 works of the Oxford ordinary language philosophers had begun to appear on the philosophy syllabus.[7] From the late 1940s throughout the 1960s, a substantial number of Australian Oxford B.Phil. graduates took up posts in the growing Australian university system. In 1965, for example, 22% of philosophers in Australia held the B.Phil. Judging by their publications, 12 of these were, for at least a brief period, practitioners of ordinary language philosophy.[8]

Austin died of cancer in 1960 at the height of his influence. After his death ordinary language philosophy rapidly declined as a force to be reckoned with at Oxford, though it continued to flourish elsewhere in Britain and in several other enclaves for another five or ten years. Austin had published very little during his relatively short life. His influence had been exerted primarily through the force of his personality and his formidable critical powers, rather than through his writings. The posthumous publication of his collected papers and two books in the early 1960s (Austin 1961, 1962, 1963) generated widespread interest in his own contribution to philosophy, but the philosophical climate within Oxford was by then already changing, largely as a result of the return to Oxford of Austin's lifelong philosophical rival, A.J. Ayer, after an absence of more than twenty years.

Ayer's avowed aim in seeking the post of Wykeham Professor of Logic in 1959 was to challenge Austin's influence.

7. Melbourne University Final Honours examination papers in philosophy are on deposit in the Special Collections Reading Room of the Baillieu Library of Melbourne University. I am grateful to Margaret Murphy, Curator of Special Collections for allowing me to photocopy the Calendar entries and examination papers.
8. Sources: *Yearbook of the Universities of the British Commonwealth*, 1967 (gives staff lists for 1965); *The Philosopher's Index*.

> I wanted to provide some local opposition to the form of linguistic philosophy which Austin had made fashionable, with what appeared to me its excessive concentration on the niceties of ordinary English usage. . .One way in which Austin had maintained his philosophical power in Oxford had been his dominance of a class ... attendance at which was confined to college tutors younger than himself. Conformably to my original intention of counter-acting his influence, I founded a rival class of a very different character (Ayer 1992).

There were a number of philosophers at Oxford who had not been sympathetic to Austin's philosophical orientation, and had resented his influence on students and his younger colleagues. Ayer's "class" encouraged the discussion of a much wider range of topics and the exploration of new developments in the more systematic, formal style of analytic philosophy then taking place elsewhere, particularly the United States. Through the efforts of Ayer and others, undergraduates and graduate students were now being challenged by a broader philosophical spectrum than was common in Austin's day. Soon enough, the examination papers began to reflect the shifting centre of gravity in Oxford's philosophical community. By the end of the decade, questions reflecting or encouraging an ordinary language approach had practically disappeared from the examination papers for either the B.A. or the B.Phil. degrees. At the same time, the composition of the teaching staff at Oxford was undergoing a significant change. Of the fifteen ordinary language philosophers present in Oxford in 1950, only six were still there in 1970. Between 1965 and 1975, Oxford awarded teaching posts to seventeen of its own B.Phil. graduates, only one of whom has published anything taking a philosophical perspective reminiscent of Austin, Ryle or the philosophers close to them in the 1950s.

Elsewhere in Britain the decline of ordinary language philosophy was less precipitous, but noticeable nevertheless. Oxford continued to dominate the job market, but fewer of its graduates entering the profession were adherents of the ordinary language approach. Sixty-six philosophers who finished the B.Phil. degree in 1965 or later held teaching posts in other British universities in 1975, of whom only ten have ever published anything that could reasonably be seen to reflect an ordinary language philosophical perspective. During this same period, the writings of the earlier group of graduates began to go in new directions, to the extent that by the end of the 1970s ordinary language philosophy had all but disappeared.

In conclusion, by adopting an epidemiological point of view, it has been possible to show that the relatively brief career of the ordinary language philosophy movement was at least in part a consequence of the fact that a distinctive new philosophical approach emerged at precisely the period in which there was a sudden and growing demand for philosophy teachers, and that it took root at the one university in Britain capable of dominating the job market, owing to its large output of philosophy graduates. That demand, at least in the early years, was not specifically for ordinary language philosophers, but simply for people qualified to teach the subject. However, the fact that a significant proportion of those who secured philosophy posts in Oxford in those early years were ordinary language philosophers ensured that a "critical mass" of like-minded philosophers were able to have a significant effect on the way philosophy was taught in the university, and this was reflected in the growing presence of "linguistic philosophy" on the examination

papers. The Oxford graduates who entered the job market in these years were those who had done very well on these examinations, and a large number of these were themselves budding ordinary language philosophers, as is evidenced by their publications. After Austin's death, and with the rapid decline of the movement within the philosophical community of Oxford, the supply of ordinary language philosophers entering the profession dried up, after about 1975, even though Oxford continued to supply the lion's share of the market. The epidemic had subsided.

Literature

Austin, J.L. (1939) "Are There *A Priori* Concepts?" *Proceedings of the Aristotelian Society*, Supplementary Volume 18, 83–105.

Austin, J.L. (1956) "A Plea for Excuses," *Proceedings of the Aristotelian Society*, 57, 1–30.

Austin, J.L. (1961) *Philosophical Papers*, G.J. Warnock & J.O. Urmson (eds.), Oxford: The Clarendon Press.

Austin, J.L. (1962) *Sense and Sensibilia*, reconstructed from the manuscript notes and edited by G.J. Warnock, Oxford: The Clarendon Press.

Austin, J.L. (1963) *How To Do Things With Words*, J.O. Urmson (ed.), Oxford: The Clarendon Press.

Ayer, A.J. (1936) *Language, Truth and Logic*, London: Gollancz.

Ayer, A.J. (1992) "My Mental Development," in L. Hahn (ed.) *The Philosophy of A.J. Ayer*, LaSalle, Ill.: Open Court.

Berlin, I. (1973) "Austin and the Early Beginnings of Oxford Philosophy," in Sir Isaiah Berlin, L.W. Forguson, D.F. Pears, G. Pitcher, J.R. Searle, P.F. Strawson, G.J. Warnock, *Essays on J.L. Austin*, Oxford: The Clarendon Press, 1–16.

Monk, R. (1990) *Ludwig Wittgenstein: The Duty of Genius*, London: Jonathan Cape.

Rogers, B. (1999) *A.J. Ayer: A Life*, London: Chatto & Windus.

Ryle, G. (1932) "Systematically Misleading Expressions," *Proceedings of the Aristotelian Society*, 32, 139–170.

Ryle, G. (1949) *The Concept of Mind*, London: Hutchinson.

The Philosopher's Index (1940–) Bowling Green State University: Philosophy Documentation Center. Dialog Information Services, Inc. (On-line Index), 1991–.

Yearbook of the Universities of the British Empire (1914–41; 1946–) London: G. Bell & Sons, Ltd. (Since 1952 it has been called *Yearbook of the Universities of the Commonwealth).*

Wittgenstein, L. (1922) *Tractatus Logico-Philosophicus*, London: Kegan Paul.

Wittgenstein, L. (1953) *Philosophical Investigations*, Oxford, Blackwell.

The Rationality of Epistemology and the Rationality of Ontology

ANDREW U. FRANK

1. Introduction

Philosophers have proposed many different ontologies. Despite hundreds of years of effort, it has been impossible to reconcile the differences between them and to establish a single, widely accepted ontology. For practical purposes a consistent and comprehensive ontology is necessary: information systems which manage adequate descriptions of the world must be constructed on the basis of some ontology, even if this ontology is never explicitly described. This was not clear in the early years of information systems and many practical problems were discovered which could later be traced back to inappropriate ontological assumptions. The connection between information systems and ontology was at the foundation of the CYC project (Lenat, Guha *et al.* 1990) and has since gained substantial acceptance among theoretical and practical thinkers in information systems (Guarino 1998; Sowa 1998). The construction of re-usable ontologies (Frank 1997) has become an interesting, rapidly growing business and 'ontologist' is an acceptable job description in forward-looking IT companies.

The design of Geographic Information Systems, which cover information about objects and properties in the world with respect to their location (Longley, Goodchild *et al.* 1999) involves ontologies too. Indeed, such systems are ontologically more demanding than ordinary administrative information systems. They span a much larger diversity of kinds of things: from the description of the elevation of the surface of the earth with a regular grid of points to the description of the natural land cover (woods, fields, etc.) and morphology (mountains, valleys, etc.). They also include man-made features like roads and buildings as well as artificial boundaries between a range of different sorts of political and administrative units (Smith 1995), etc. There is no ready-made single ontology to cover all of these most diverse aspects of reality. Therefore we propose here the construction of an ontology consisting of several coordinated tiers.

An ontology constructed from tiers can integrate different ontological approaches in a unified system. In particular, it can merge a plenum, continuous space ontology with Aristotle's 'natural kind' ontology of objects. We can also integrate the ontology of 'social reality' described by Searle (1995). It seems possible also to overcome some of the differences between competing proposals, differences which we can understand as motivated by the examples the authors have in mind. From our practical experience, we have learned that a single ontology, which applies to all situations and the most diverse kinds of phenomena in the world or in our imagination, is not achievable. Therefore we propose here an orderly integration of otherwise contradictory proposals.

I am not interested here in terminological discussions, and I use terms like 'ontology'

B. Brogaard, B. Smith (Hg.), *Rationality and Irrationality / Rationalität und Irrationalität*, S. 110–119.

in a generic way; Guarino (1997) has shown the many different uses of the term by different authors and I do not want to add to this list. My approach is empirical and stresses our daily experience in interacting with the world as a source of knowledge to build ontologies. The goal is a computational model of an ontology, which can be used for the construction of information systems.

The remainder of this paper first gives an overview of the tiers and then discusses each of them in turn. It sketches how a computational model of the ontology could be built and draws some conclusions about its usefulness.

2. The Five Tiers of the Ontology

An ontology for an information system is necessarily based on a realist position. Therefore tier 0 of the ontology assumes that there exists a physical reality, which may best be imagined as a four-dimensional continuous field of attribute values. This could be looked at as the ontology proper, where the next tiers are perhaps more similar to what some authors would assign to the realm of epistemology. Tier 1 covers the point-wise observation of this reality by cognitive agents. Tier 2 discusses how agents form objects from point-wise observations; this is somewhat similar to Aristotle's metaphysics. Tier 3 embraces social reality in the sense of Searle (1995) and other similar socially constructed elements (Berger and Luckmann 1996). Tier 4, finally, deals with the ideas cognitive agents have about the world.

Tier O:	human-independent reality
Tier E1:	observation of physical world
Tier E2:	objects with properties
Tier E3:	social reality
Tier E4:	subjective knowledge

Fig. 1. The five tiers of ontology

The discussion here excludes the effects of learning on the ontology; it describes what is true when we consider a short period (days, weeks) and excludes the changes which are possible through extended experiences in an environment.

3. Physical Reality Seen as an Ontology of a Four-Dimensional Field

The physical laws which describe the behavior of the macroscopic world can be expressed as differential equations, which describe the interaction of a number of properties in space – the whole seen as forming a continuum. For each point in space and time a number of properties can be observed: color, the forces acting at that point, the material and its properties (like mass, melting temperature at the point, and so on). Movement of objects can be described as changes in these properties; even the movement of solid ob-

jects can be described, the cohesive forces in the body maintaining its shape. The description of reality via differential equations (e.g., the description of forces in a plate under a load) is widely used in mechanical and civil engineering, geology, etc. This view is also quite natural for most 'global systems' studies (Mounsey and Tomlinson 1988).

A field can be observed at every point in space and time for different properties:

$$f(x, y, z, t) = a.$$

Abstracting from the temporal effects, a snapshot of the world can be described by the formula which Goodchild called 'geographic reality' (Goodchild 1992).

$$f(x,y,z) = a$$

The processes occurring in this physical reality have spatial and temporal extensions: some are purely local and happen very fast; others are very slow and affect very large regions. The processes of objects moving on the tabletop are fast (m/sec) and the spatial extent is small (m); movement of persons in cities is again fast (m/sec) and the movements of the buildings very slow (mm/annum); geological processes are very slow (mm/annum) and affect large areas (1000 km^2). One can thus associate different processes with different frequencies in space and time (Fraser 1981). Each science has a certain scope: it is concerned with processes in a specific spectrum of space and time which interact strongly; other processes, not included in this scope, appear then to be either so slow or so fast that they can be considered constant.

Space and time form together a four-dimensional space in which other properties are organized. Giving space and time a special treatment results in simpler formulations of the physical laws that are of particular interest to humans. For example, the mechanics of solid bodies, e.g., the movement of objects on a tabletop, is explainable by Newtonian mechanical laws, which relate phenomena which are easily observable for humans in a simple form ($s = v\,t$, etc.). Other sciences, for example, astrophysics, prefer other coordinate systems in which mass is included.

However, the assumption that the formula $a = f(x,y,z,t)$ describes a regular function in the sense of a function which yields only one single value is equivalent to the assumption that there is only one single space-time world and excludes 'parallel universes' as parts of reality.

4. Observation of Physical Reality

Agents can – with their senses or with technical instruments – observe the physical reality at the current time, the 'now'. Results of observations are measurement values on some measurement scale (Stevens 1946), which may be quantitative or qualitative.

Observation with a technical measurement system such as remote sensing comes very close to be an objective, human-independent observation of reality. A subset of the phenomena in reality is objectively observed. Many technical systems allow the synchronous observations of an extent of space at the same time, for example, remote sensing of geographic space from satellite. A regular grid is used and the properties observed are energy reflected in some bands of wavelength (typically the visible spectrum plus some part of infrared).

Observation through sampling of many points is effected also by our eyes, but it is also used by robots, where TV cameras which sample the field in a regular grid are used to construct 'vision' systems to guide the robot's actions in manipulating objects or guiding the robot's movements through buildings (Kuipers 1998).

Observations of reality are always marked by imprecision – the knowledge we acquire is never perfect. The technical effects of our measurement systems allow us at best measurements up to 10^{-13}, which is, incidentally, much worse than the theoretical limits imposed by the Heisenberg uncertainty principle.

5. Objects with Properties

Our cognitive system is so effective because, from the array of sensed values, it forms individuals, which are usually called objects, and it reasons about them. Thinking of tables and books and people is much more effective than seeing the world as consisting of data values for sets of cells, regularly subdivided across a grid (i.e., three-dimensional cells, often called voxels). It is economical to store properties of objects and not deal with individual raster cells. As John McCarthy and Patrick Hayes have pointed out:

> ... suppose a pair of Martians observe the situation in a room. One Martian analyzes it as a collection of interacting people as we do, but the second Martian groups all the heads together into one subautomaton and all the bodies into another. ...How is the first Martian to convince the second that his representation is to be preferred? ...he would argue that the interaction between the head and the body of the same person is closer than the interaction between the different heads. ... when the meeting is over, the heads will stop interacting with each other but will continue to interact with their respective bodies. (McCarthy and Hayes 1969, p. 33)

Our experience in interacting with the world has taught us that the most appropriate subdivision of continuous reality is that into individuals. The latter are most often continuous in space and endure in time. Instead of reasoning with arrays of connected cells, as is done, for example, in computer simulations of strain analysis or oil spill movements, we select the more economical and more direct mode of reasoning with individuals: The array on the tabletop is divided into objects at the boundaries where cohesion between cells is low; a spoon consists of all the material which moves with the object when I pick it up and move it to a different location. This is obviously more effective than individual efforts to reason about the content of each cell. In an ever changing world, objects are typically formed in such a way that many of their properties remain invariant over time, which further simplifies reasoning. Animals and most plants form individuals in a natural way.

The cognitive system operates very quickly in identifying objects with respect to typical interactions. We see things as chairs or cups if they are presented in situations where

sitting or drinking are of potential interest. Under other circumstances, the same physical objects may be seen as a box and a vase. The detection of 'affordances' of objects is immediate and not a product of conscious reasoning. The identification of affordances implies a breakup of the world into objects: the objects are what we can interact with (Gibson 1979).

Cognitive science has demonstrated that small infants as early as three months have a tendency to group what they observe in terms of objects and to reason in terms of objects. It has been shown that animals do the same. Most of the efforts of our cognitive system to structure the world into objects are unconscious and so it is not possible for us to scrutinize them. There are a number of well-known effects where the same image is interpreted in different ways, for example, the well known Necker-cubes which can be seen as cube or a corner, but not both at once. But such examples are rare. The default process assigns objects univocally.

Efforts to explain the categorization of phenomena in terms of common nouns based on a fixed set of properties were initiated by Aristotle. These occasionally lead to contradictions. Dogs are often specified as 'can bark', 'have four legs', etc., but from such a set of attributes it does not follow that my neighbor's dog, which lost a leg in an accident, is no longer a dog. Modern linguistics and psychology assume generally that prototype effects make some exemplars better examples of a class than others. A robin is a better example of a bird than is a penguin or an ostrich (Rosch 1973; Rosch 1978). Linguistic analysis suggests that the ways objects are structured are closely related to the operations one can perform with them, and empirical data support this (Jackendoff 1983; Fellbaum 1998).

Humans have a limited set of interactions with the environment – five senses to perceive it and operations like walking, picking up, etc. – and these operations are common to all humans. Therefore the object structure – at least at the level of direct interaction - is common to all humans and it provides the foundation on which to build the semantics of common terms (Lakoff 1988). In general, the way individual objects and object types are formed varies with the context, but is not arbitrary. This commonality in the basic experiences of all humans gives sufficient grounding for the semantics of everyday words.

6. Social Ontology

Human beings are social animals; language allows us to communicate and to achieve high levels of social organization and division of labor. These social institutions are stable, evolve slowly and are not strongly observer dependent. Conventionally fixed names for objects, but also much more complex arrangements which are partially modeled according to biological properties, for example, the kin system (Lévi-Strauss 1967), or property rights derived from physical possession, can be refined and elaborated to the complex legal system of today's society.

6.1 Names

The types of proper and common names used in our various natural languages are clearly the result of a social process: proper names are words used for individuals, which identify

objects in ways which are different from predicates to select individuals based on unique sets of properties. Such socially agreed identifiers seem to be a property of the individual, because they exist outside of the observing agent. Pointing out that 'chien', 'Hund' and 'cane' are equally good words to describe what in English is called a dog, should make it clear that none of these names is more natural than any other. Examples for proper names and similar identifiers reach from names for persons and cities to license plates for cars; there are also short-lived names created, like 'my fork', during a single dinner.

6.2 Institutions

Social systems construct rules for their internal organization (Berger and Luckmann 1996), for example, laws, rules of conduct and manners, ethics, etc. Such rules are not only procedural ("thou shalt not kill"), but often create new conceptual objects (e.g., marriage in contradistinction to cohabitation without social status), adult person (as a legal definition and not a biological criterion), and so on. Institutions are extremely important in our daily life and appear to us as real; who would deny the reality of companies, such as the Microsoft Corporation?

Much of what administration and therefore administrative databases deal with are facts of law – the classification of reality in terms of the categories of the law. The ontology of these objects is defined by the legal system and is only loosely related to the ontology of physical objects; for example, legal parcels behave in some ways similar to liquids: one can merge them, but it is not possible to recreate the exact same parcels again (without the agreement of the mortgage holders) (Medak 1999; Medak in press).

7. Ontology of Cognitive Agents

Cognitive agents – persons and organizations – have incomplete and partial knowledge of reality, but they use this knowledge to deduce other facts and they make decisions based on such deductions. Agents are aware of the limitations of the knowledge of other agents; social games, social interaction and business are to a very large degree based on the reciprocal limitations of knowledge. Game theory explores rules for behavior under conditions of incomplete knowledge (von Neumann and Morgenstern 1944; Davis 1983; Baird, Gertner *et al.* 1994).

The knowledge possessed by a person or an organization increases over time, but the knowledge lags necessarily behind the changes on the side of reality. Decisions are made based on this not quite up-to-date knowledge. Fairness dictates that the actions of agents are judged not with respect to perfect knowledge but rather with respect to the incomplete knowledge the agent had or should have had if he had shown due diligence. Sometimes the law protects persons who have no knowledge of certain facts. The popular saying is "Hindsight is 20/20" or "afterwards, everybody is wiser". A fundamental aspect of modern administration is the concept of an audit: administrative acts must be open to inspection so that it can be established whether they were performed according to the rules and regulations. Audits must be based on the knowledge available to the agent, not on the facts discovered later. For audits it must therefore be possible to reconstruct the knowledge

which an agent, for example, in public administration, had at a certain time. This leads to the bi-temporal perspectives usually differentiated in a database: the time a fact becomes true in the world and the time the agent acquires knowledge of this fact (Snodgrass 1992).

8. Computational Model of a Tiered Ontology

The design of the tiered ontology is oriented towards the construction of a computational model. The demonstration of misunderstandings and terminological difficulties in various texts on ontology but also the observation of problems in practice with differences in the interpretation of terms have led us to investigate computational models which reduce our reliance on natural language terminology. Algebras define terms up to an isomorphism without regress to other, previously defined terms, which is exactly what is necessary to define the behavior of objects in reality or their simulated behavior in an information system. Between reality and information system we should have as far as possible an isomorphism. The two realms – reality and information – are connected by the experience of the agent interacting with the world based on his knowledge.

Certain parts of the ontology have been translated into computational models in a multi-agent setting (Weiss 1999). Multi-agent systems, the way we use them, are systems in which we simulate agents, including their bodies and perceptual and cognitive systems, in a simulated reality. We have completed one such simulation in which one agent explores a simplified city and then draws a map, which is later used by another agent to navigate (Figure 2). We have also completed a simulation for social reality (Bittner in progress) wherein the meanings of terms like 'ownership' and 'land' are defined. Agents then follow the rules of real estate law in dealing with the simulation. It seems possible to construct a computational model of the complete five tiers of the ontology in this framework.

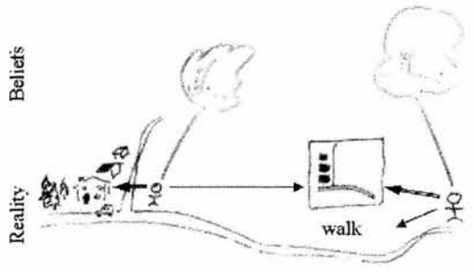


Fig.2. An agent producing a map and another agent using a map for navigation

We found it extremely useful to have a way of formally checking that the descriptions are complete, i.e., that all parts which are used to define a concept are in turn defined

somewhere else in terms of a very simple set of primitives. Checking that the types of inputs and outputs correspond to each other – something which can be done automatically – gives additional confidence that the model is logically consistent (Milner 1978; Jones 1994). Running the computational model allows us finally to test whether the model reflects the intended behavior correctly. We found the public domain functional language Haskell (Peterson, Hammond *et al.* 1996) extremely useful for this purpose.

9. Conclusion

In today's world of networked information systems, the clarification of the ontological bases used to collect and manage data becomes ever more important. Questions of interoperability (Goodchild, Egenhofer *et al.* 1998) are very often essentially ontological questions. In this environment practical ontologies – ontologies which work – become necessary. They can help us to understand how to integrate data from different sources, and possibly in a single system. This topic will be further explored in the REVIGIS project (REVIGIS 2000).

We have sketched here a program of a tiered ontology, where different approaches are used on each tier. We follow an empirical approach, and integrate different ways of forming an ontology to achieve a practically useful solution. Our experiments so far suggest that computational models for ontology are possible. This would be a substantial step towards practically useful ontologies for information systems.

This text is a brief version of a much more detailed description which will be published in a book from the EU project Chorochronos, where it provides the ontology for the design of spatio-temporal databases (Sellis and Koubarakis to appear).

10. Acknowledgments

The work by Barry Smith and Nicola Guarino on ontologies has influenced my thinking in many details. I thank Werner Kuhn, Max Egenhofer, Annette von Wolff, Martin Raubal and Christine Rottenbacher for the time they have taken to discuss these ideas with me. Roswitha Markwart has improved the text of this contribution.

Support from the Chorochronos and REVIGIS project (both European Commission) and a project on formal ontology for land registration systems (financed by the Austrian Science Fund) are gratefully acknowledged.

References

Baird, D. G., R. H. Gertner, et al. (1994). *Game Theory and the Law*. Cambridge, Mass., Harvard University Press.

Berger, P. L. and T. Luckmann (1996). *The Social Construction of Reality*. New York, Doubleday.

Bittner, S. (in progress). *An agent-based model of reality in a cadastre*. Ph.D. thesis, De-

partment of Geoinformation. Vienna, Technical University Vienna.
Davis, M. D. (1983). *Game Theory*. Minneola, NY, Dover Publications.
Fellbaum, C., Ed. (1998). *WordNet: An Electronic Lexical Database*. Language, Speech, and Communication. Cambridge, Mass., The MIT Press.
Frank, A. U. (1997). *Spatial Ontology: A Geographical Information Point of View*. Spatial and Temporal Reasoning. O. Stock. Dordrecht, Kluwer: 135–153.
Fraser, J. T., Ed. (1981). *The Voices of Time*. Amherst, The University of Massachusetts Press.
Gibson, J. (1979). *The ecological approach to visual perception*. Hillsdale, NJ, Erlbaum.
Goodchild, M. F. (1992). "Geographical Data Modeling." Computers and Geosciences 18(4): 401–408.
Goodchild, M. F., M. Egenhofer, et al., Eds. (1998). *Interoperating Geographic Information Systems* (Proceedings of Interop '97, Santa Barbara, CA). Norwell, MA, Kluwer.
Guarino, N. (1997). "Understanding, building, and using ontologies: A commentary to "Using Explicit Ontologies in KBS Development", by van Heijst, Schreiber, and Wielinga." *International Journal of Human and Computer Studies* 46: 293–310.
Guarino, N. (1998). *Formal Ontology and Information Systems. Formal Ontology in Information Systems* (Proceedings of FOIS98, Trento, Italy, 6–8 June, 1998). N. Guarino. Amsterdam, IOS Press: 3-15.
Jackendoff, R. (1983). *Semantics and Cognition*. Cambridge, Mass., MIT Press.
Jones, M. P. (1994). *Qualified Types: Theory and Practice*, Cambridge University Press.
Kuipers, B. (1998). *A hierarchy of qualitative representations for space. Spatial Cognition – An Interdisciplinary Approach to Representing and Processing Spatial Knowledge*. C. Freksa, C. Habel and K. F. Wender. Berlin Heidelberg, Springer-Verlag. 1404: 337–350.
Lakoff, G. (1988). *Cognitive Semantics. Meaning and Mental Representations*. U. Eco, M. Santambrogio and P. Violi. Bloomington, Indiana University Press: 119–154.
Lenat, D. G., R. V. Guha, et al. (1990). "CYC: Toward programs with common sense." *Communications of the ACM* 33(8): 30–49.
Lévi-Strauss, C. (1967). *Structural Anthropology*, Basic Books.
Longley, P., M. Goodchild, et al., Eds. (1999). *Geographical Information Systems* – Volume 1: *Principles and Technical Issues*; Volume 2: *Management Issues and Applications*. New York, John Wiley & Sons.
McCarthy, J. and P. J. Hayes (1969). *Some Philosophical Problems from the Standpoint of Artificial Intelligence. Machine Intelligence 4*. B. Meltzer and D. Michie. Edinburgh, Edinburgh University Press: 463–502.
Medak, D. (1999). *Lifestyles – A Paradigm for the Description of Spatiotemporal Databases*. Ph.D. thesis, Department of Geoinformation. Vienna, Technical University Vienna.
Medak, D. (in press). *Lifestyles. Life and Motion of Socio-Economic Units*. A. U. Frank, J. Raper and J.-P. Cheylan. London, Taylor & Francis.
Milner, R. (1978). "A Theory of Type Polymorphism in Programming." *Journal of Computer and System Sciences* 17: 348–375.
Mounsey, H. and R. F. Tomlinson, Eds. (1988). *Building Databases for Global Science – Proceedings of the IGU Global Database Planning Project, Tylney Hall, Hampshire, UK, 9–13 May 1988*. London, Taylor & Francis.

Neumann von, J. and O. Morgenstern (1944). *Theory of Games and Economic Behavior*. Princeton, NJ, Princeton University Press.
Peterson, J., K. Hammond, et al. (1996). Report on the functional programming language Haskell, Version 1.3. http://www.haskell.org – Research Report YALEU/DCS/RR–1106, Yale University.
REVIGIS (2000). The REVIGIS Project Web Page. URL: http://www.cmi.univ-mrs.fr/REVIGIS/.
Rosch, E. (1973). "Natural categories." *Cognitive Psychology* 4: 328–350.
Rosch, E. (1978). *Principles of Categorization. Cognition and Categorization*. E. Rosch and B. B. Lloyd. Hillsdale, NJ, Erlbaum.
Searle, J. R. (1995). *The Construction of Social Reality*. New York, The Free Press.
Sellis, T. and M. Koubarakis, Eds. (to appear). *Spatio-Temporal Databases*. Berlin Heidelberg, Springer-Verlag.
Smith, B. (1995). *On drawing lines on a map. Spatial Information Theory – A Theoretical Basis for GIS (Int. Conference COSIT'95)*. A. U. Frank and W. Kuhn. Berlin Heidelberg, Springer-Verlag. 988: 475–484.
Snodgrass, R. T. (1992). *Temporal Databases. Theories and Methods of Spatio-Temporal Reasoning in Geographic Space (Int. Conference GIS – From Space to Territory, Pisa, Italy)*. A.U. Frank, I. Campari and U. Formentini. Berlin, Springer-Verlag. 639: 22–64.
Sowa, J. F. (1998). *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. Boston, PWS Publishing.
Stevens, S. S. (1946). "On the theory of scales of measurement." *Science* 103(2684): 677–680.
Weiss, G. (1999). *Multi-Agent Systems: A Modern Approach to Distributed Artificial Intelligence*. Cambridge, Mass., The MIT Press.

Reason and Necessity

NEWTON GARVER

1. Preliminary Remarks

Necessity is a constitutive element of reason. To give the reason for some action or event – to explain it rationally – is to show why it had to happen or why it must happen. To say this much is not to embark on risky or obtuse philosophy. This much can be gathered simply by looking at common practices of explaining and rationalizing events and actions. Galileo's Law provides a major premise from which to conclude that apples, cannonballs, ballistic missiles, and other objects *must* move as they do move. That is how it furnishes an explanation. Except for the necessity, we would have only a succession of events and no explanation, *ante hoc ergo propter hoc* is a fallacy within the domain of rational thought. Rational justification in practical reason follows a similar pattern, although the argumentation is often less apparent because the major premise often remains suppressed. It is normally acceptable behavior to choose from a number of available options, but my behavior remains without rational justification unless I show that all the other options are morally or rationally unacceptable. If there are principles from which I can deduce that the other alternatives are all unacceptable, I am left with the one action which I *must* perform. That is why George Bush claimed at the beginning of the Gulf War that Saddam Hussein left him no choice, and why Bill Clinton said the same thing about Slobodan Milosevic when he ordered the bombing of Kosovo. Unless something is necessary – that is, if there would have been other possible outcomes or alternative possible actions – what happened or will happen has not yet fallen with in the grasp of reason. Theoretical reason seeks to establish the necessity of an event, given laws and antecedent conditions. Practical reason seeks to establish the necessity of an action, given antecedent obligations and circumstances.

Wittgenstein had little use for either theoretical or practical reason. He had, furthermore, a disdain for the pretensions of reason. Although giving rational explanations is a common "language-game" that very likely transcends cultural boundaries and that appears in pre-school children, it is not one that Wittgenstein endorses. He rarely even speaks about it, although his sharp rebuke to Fraser's *Golden Bough* is in part a critique of rational explanation gone astray. It is partly this silence that leads me to think that he disdains reason, partly his frequent practice of noting alternatives or intermediary cases, partly his explicit rejection of explanation in philosophy (PI 126, 654–655), and partly some terse remarks in the *Tractatus* that seem to me to underlie the other considerations.

What he says in the *Tractatus* is:

> There is no compulsion making one thing happen because another has happened. The only necessity that exists is *logical* necessity. (TLP 6.37)

B. Brogaard, B. Smith (Hg.), *Rationality and Irrationality / Rationalität und Irrationalität*, S. 120–128.

> Just as the only necessity that exists is *logical* necessity, so too the only impossibility that exists is *logical* impossibility. (6.375)

There is nothing about "existence" in the German. TLP 6.375, for example, reads:

> Wie es nur eine *logische* Notwendigkeit gibt, so gibt es auch nur eine *logische* Unmögligkeit.

I will therefore make use of a simpler English translation:

> There is only *logical* necessity ... Just as there is only *logical* necessity, so too there is only *logical* impossibility.

These remarks in the *Tractatus* demonstrate unequivocally the influence of Hume on Wittgenstein's thought. To many readers they seem false, because they fly in the face of our everyday experiences of necessity and impossibility. That is so because they undermine our conventional conceptions of both science and ethics. Their context in the *Tractatus* is Wittgenstein's discussion of science and scientific method. They come near the end of the discussion, as a kind of summary or epiphany, and they immediately precede his discussion in the 6.4's of ethics and value. If we sense that these remarks undermine our conventional conception of science and ethics (hence of theoretical reason and of practical reason), the context should make it clear that this is precisely what Wittgenstein intended. In this paper I wish to consider the metaphysical basis for Wittgenstein's striking remarks, show how they are transformed and retained in his later work, and then review their implications for his (and our) sense of morality.

2. Metaphysics

The metaphysics of the *Tractatus* appears radically dualistic. We should nonetheless bear in mind that little about the *Tractatus* is just what it initially seems. Wittgenstein begins by remarking that the world, reality, is the totality of facts. Facts are contingent. Not only might each fact be different, or not occur, but it is also contingent that the world exists; that is, that there are any facts at all. In the world as Wittgenstein sketches it at the beginning of the *Tractatus* there is no room for necessity or impossibility. Wittgenstein supplements this Humean metaphysics almost immediately through the introduction of two further categories of thought, *substance* and *language*. These two categories are related; they are not independent of one another. Of the two, it is language that is dominant. Not only is Wittgenstein's main focus on describing or explaining the essence of language (see McGuinness 1988, chapter 9), but one also comes upon *language* before *substance* if one reads through the whole numbers and single decimals of the work, as Wittgenstein (in his only footnote) implies we ought to do in order to see what is most important. 2.1 reads: "We picture facts to ourselves." This picturing is elementary linguistic activity. With this remark humans and human activity come into play. This human activity of picturing facts to ourselves is a wholly contingent feature of our world. That we do this sort

of thing is a contingent fact, like all others: the world might well have existed without humans and human language. From 2.1 on, the *Tractatus* never further considers how the world is in itself, independently of its facts being pictured.

Just prior to 2.1, in the 2.0's (with that seemingly meaningless decimal inserted), Wittgenstein presents some powerful and puzzling ideas about the substance of the world. Unlike the world, which consists "of facts, not of things" (TLP 1.1), the substance of the world consists of objects (*Gegenstände*), thereby introducing what appears to be a sort of metaphysical dualism, for there is no clear way for "objects" to be fully grounded in "facts" nor for "facts" to be fully grounded in "objects". That the doctrine of substance is inserted as a bridge between the world of facts and our picturing of facts is obvious from the text and is not a matter of controversy. What remains a matter of controversy is whether or not the *Gegenstände* are, like facts, elements of reality. Malcolm (1986) and Pears (1987) believe that they are, thereby exacerbating the apparent embarrassment of metaphysical dualism. But it seems wiser to follow the suggestion of Erik Stenius (1960), namely to read many of the longer decimals as partly anticipatory of what is to come rather than merely as comments and elaborations on preceding remarks. Admittedly such a reading of the *Tractatus* flies in the face of Wittgenstein's own explanation of the decimals. Nonetheless, McGuinness (1981) and Ishiguro (1969), as I have argued elsewhere (Garver 1994, chapter 6), make a convincing case for excluding substance from the realistic spirit of the *Tractatus*, a case which has subsequently been reinforced by Kenny (1984), Diamond (1991), Conant, Goldfarb, and others. Following these latter suggestions, the doctrine of substance is not to be seen as a comment on the world of facts but rather as a comment on the presuppositions of our picturing of facts.

Such a reading not only reduces the embarrassment of metaphysical dualism but also helps to underline the centrality of language (picturing) as the main focus of the *Tractatus*. Necessity and impossibility do not reside in the world or in the nature of facts, which are one and all contingent, but rather in unavoidable features of our picturing of the world. Perhaps the most dramatic and eloquent statement of this relationship occurs in TLP 5.511, following Wittgenstein's introduction of some further logical symbolism:

> How can logic, all-embracing logic which mirrors the world, use such peculiar crotchets and contrivances? Only because they are all connected with one another in an infinitely fine network, the great mirror.

Substance, consisting of simple objects, has the Kantian status of being a transcendental requirement of propositions, that is, of our making ourselves pictures of facts. Both substance and propositions further presuppose logic. In the presentation of his doctrine of substance Wittgenstein had written (2.012), "In logic nothing is accidental: if a thing *can* occur in a state of affairs, the possibility of the state of affairs must be written into (*präjudiziert in*) the thing itself." In the first pages of the *Tractatus*, therefore, Wittgenstein makes clear that he sees logic connected with characteristics of substance rather than with the world of facts. That propositions, or truth-claims, presuppose logic is perhaps the main insight of the *Tractatus*. One expression of it occurs in TLP 3.42: "A proposition determines only one place in logical space: nevertheless the hole of logical space must already be given by it." But it is clear that nothing about this indispensability

of logic entails that logic is a feature of reality. That logic does not belong to the world of reality follows directly from Wittgenstein's "fundamental idea": "My fundamental idea is that the 'logical constants' are not representatives [*nicht vertreten* – do not stand for anything]; that there can be no representatives of the *logic* of facts." This contrast between logic and reality is expressed again in TLP 4.12, which introduces the discussion of logic, formal concepts, internal relations, and truth tables:

> Propositions can represent (*darstellen*) the whole of reality, but they cannot represent what they must have in common with reality in order to be able to represent it – logical form.

If logical form does not belong to the world of reality, then the objects that make up the substance of the world do not belong to that world either. Wittgenstein goes on to say that the logical features of our representations of facts can be *shown*, adding that "What *can* be shown, *cannot* be said" (TLP 4.1212).

This reading of the *Tractatus* helps not only to see the continuity in Wittgenstein's work but also to understand an important part of what he took from Kant. Kant's "Copernican Revolution" consists in insisting that some constitutive features of our experience derive from our ways of responding to the world rather than from the world itself. He was led to this insight at least in part, as al-Azm (1972) has shown, through deep and troubled consideration of the controversy between Leibniz and Clarke about key concepts in Newtonian mechanics – about space in particular, but also about time and causation. Kant was struck by the force and cogency of the negative arguments of each contestant, and in his own work he came to accept those negative arguments as decisive. But the positions of Leibniz and Clarke seemed exhaustive as well as exclusive alternatives – both parties agreed about that. Kant saw that the only way out of the dilemma, if the negative arguments on both sides were sound, was to reject what both parties agreed about, by positing another alternative. The result is Kant's subtle diagnosis of "transcendental illusion" as conceiving space, time, and causation to be constitutive features of reality itself rather than as regulative features of our rational response to reality. The transcendental illusion which is found in the cosmological theories of both Clarke and Leibniz, he decided, results from making constitutive (ontological) use of regulative concepts. Wittgenstein is no doubt following Kant in regarding substance, objects, propositions, logic, and necessity as features of our making pictures of facts (regulative in Kant's sense), rather than as features of facts or of reality. He gives a fresh endorsement of this Kantian perspective in PI 104:

> We predicate of the thing what lies in the method of representing it. Impressed by the possibility of a comparison, we think we are perceiving a state of affairs [*Sachlage*] of the highest generality.

3. Transformation

I have presented a view of the *Tractatus* that minimizes the differences between the early

work and Wittgenstein's later work. There are, of course, major differences. Two of the principal ones are that the *Investigations* recognizes a very large number of "language-games" (or uses of language) in addition to picturing facts; and (Garver 1999) that the work also acknowledges that vagueness is unavoidable and therefore develops the radically different view that clarity depends on perspicuous representation rather than on analysis. Many commentators, beginning with Russell and well represented by Gellner (1992), believe that the rigor of Wittgenstein's earlier work does not and cannot survive these changes, either because ordinary language is too vague or because one can always switch to another language-game that is less rigorous. Such commentators are misguided, and it will be worthwhile to sketch the reasons why.

One can appreciate the rigor and philosophical sting of Wittgenstein's later work by considering what he has to say about knowledge-claims. His best-known remark on this topic occurs in the midst of his discussion of the possibility of private language:

> It can't be said of me at all (except perhaps as a joke) that I *know* that I am in pain. What is it supposed to mean – except perhaps that I *am* in pain? (PI 246)

Of course it cannot be said of me either that I *don't* know that I am in pain. No one can seriously suppose that uncertainty is implied. The point is rather the reaffirmation of a tenet he held in common with the Vienna Circle, that *knowledge* belongs to the domain of empirical science. So where neither doubt nor empirical evidence come into play (as is that case with my claim to be in pain), whatever is said cannot be knowledge-claim. Wittgenstein differs from some empiricists in insisting that some truth-claims are not knowledge-claims; but he thereby ends up with a tougher, sharper conception of knowledge.

Throughout his later work Wittgenstein contrasts empirical propositions from both avowals and grammatical propositions. One can have *knowledge* only of empirical propositions, not of grammatical ones or avowals, partly because one can *know* something only if others can also know it on the same evidentiary basis. Philosophical confusion is likely to arise from failure to appreciate these distinctions, and some of it again surrounds necessity. In this later framework, necessity would be a feature of grammar rather than just of logic. One interesting thing about the *Investigations*, however, is that Wittgenstein rarely speaks of necessity there, and never with the clarity and directness of TLP 6.37. The one explicit use of the word occurs in PI 372:

> Consider: "The only correlate in language to an intrinsic necessity [*Naturnotwendigkeit*] is an arbitrary rule. It is the only thing which one can milk out [*abziehen*] of this intrinsic necessity into a proposition."

This is not a passage in which Wittgenstein is speaking forthrightly in his own voice. The prefatory injunction and the quotation marks signify both distance from and caution about this intriguing idea. One danger is that this idea might well lead one to invoke what might be called "special logics" or "parochial grammars." That is, one might establish, or try to establish, arbitrary rules for a particular discourse and attempt to defend spurious claims of necessity – e.g. those of Bush and Clinton – in this manner. If successful, this

sort of exculpation would open the floodgates to moral and linguistic laxity. Such a prospect is no doubt one of the reasons for Wittgenstein's distance and caution with respect tot he idea presented in PI 372.

One can obtain some insight about Wittgenstein's perspective on these matters by considering his sharp rebuke to Norman Malcolm about "national character" in 1939. There was a report of the Germans accusing the British of having tried to assassinate Hitler with a bomb, and Malcolm, by his own account, had said that "such an act would be incompatible with the British 'national character'." Malcolm goes on: " My remark made Wittgenstein extremely angry. He considered it a great stupidity and also an indication that I was not learning anything from the philosophical training he was trying to give me." (Malcolm 1984, p.30) Wittgenstein's rebuke shows decisively that he would have nothing to do with such "special logics." Rather, he continued to hold to Frege's view that there is just one logic, the same for all languages there have been or will be. I cannot here show the evidence for the enduring influence of Frege on Wittgenstein's thought in this respect, but Cora Diamond argued it in convincing detail in her address to the American Philosophical Association in December 1999.

In Wittgenstein's later work it remains the case that necessity and impossibility, though rarely mentioned, are not part of the natural world, the reality with which we must cope. If one looks at the contexts where the word 'must' occurs in the English translation of the *Philosophical Investigations*, one finds numerous cautions against thinking that something *must* be the case. Typical of these passages is PI 66:

> Consider for example the proceedings that we call "games". I mean board-games, card-games, Olympic games, and so on. What is common to them all? – Don't say: "There *must* be something common, or they would not be called 'games'" – but *look and see* whether there is anything common to them all.

Wittgenstein in his later work continues to give no support to logical realists (who hold necessity and impossibility to reside in objective reality), and he refuses to succor cultural or moral relativists (who see necessities and impossibilities introduced by special logics or parochial grammars). Instead he repeatedly reminds us of empirical realities, intermediate cases, and alternatives possibilities.

4. Morals

Wittgenstein's undermining of practical reason does not weaken the role of morality, any more than his undermining of theoretical reason weakens the role of empirical science. His rebuke to Malcolm is an instructive instance of his strict moralism. Malcolm's misstep was both a logical and a moral failure. By supposing that certain acts would be impossible for British agents he was guilty of a logical or grammatical error, for "there is only *logical* impossibility." Malcolm's moral mistake was partly his failure to make everyday use of his study of philosophy – Wittgenstein was surely stricter in this regard than most of us. Malcolm, however, was also making a morally dubious moral judgment, one which cleared certain people of possible moral fault on the basis of an "impossibil-

ity"; that is, not on the basis of evidence but on the basis of a refusal to consider evidence. For the words "incompatible with British national character" here function just as the "must" in PI 66, to block out the empirical evidence.

Necessity and impossibility are often invoked in moral discourse, sometimes to justify force or violence, as in the claims of Bush and Clinton, but always to exculpate by canceling moral responsibility. Given the familiar saying, "Ought implies can," specific claims of necessity or impossibility are presented so as to remove an action from the scope of moral judgment. Such a move might be conceived as "logical" in a spurious sense, since, if successful, it would change the rules of the discourse. Wittgenstein's remarks in the *Tractatus* (6.37–6.375) cut short such escapes from moral responsibility. There are no escapes from responsibility, according to the *Tractatus*. Since there is only *logical* necessity or impossibility, there is always freedom and the responsibility that goes with it. This is the moral toughness in Wittgenstein's position — a toughness echoed decades later, without acknowledgement, by Jean-Paul Sartre's doctrine that any denial of freedom or responsibility (notably by pleading necessity or impossibility) is an act of bad faith.

One should bear in mind that the critical impact of Wittgenstein remarks falls squarely on a widespread but rather crude popular version of practical reason. If practical reason is articulated in terms of obligations and ideals, rather than laws and necessities, as Onora O'Neill (1989) proposes, it escapes the thrust of Wittgenstein's comments. O'Neill's version of practical reason at the same time sharpens rather than blunts the requirements and expectations of moral responsibility, which is in line with Wittgenstein's remark.

5. Conclusion

I find Wittgenstein's tough morality exciting and challenging but unsatisfactory. Part of the dissatisfaction comes from considering Wittgenstein's own life. While he cast off many social "necessities" (customs, conventions) and so did things that others thought impossible, he also seems to have set himself other necessities and impossibilities. He had to write down his thoughts, he had to seek clarity in his thoughts and in the expression of his thoughts, he could be satisfied with his achievements, he could not be frivolous, and so on. In this pattern of necessities and impossibilities, as well as in his rejection of conventional "necessities", we see his character and identity emerge. So it is for most of us. These impossibilities and necessities are contingent rather than absolute, in line with the *Tractatus*. But they are intentional rather than simply empirically contingent. They enhance rather than negate responsibility. It is difficult to imagine living without them.

Often I am unclear about what to do, but sometimes I have a clear sense that something is necessary or impossible for me. I must finish this paper. Whereas Wittgenstein in his later years found it impossible to submit a paper for publication, I often find it necessary. Among the impossibilities I am committed to are military service and certain loyalty oaths. In the case of military service, I found it impossible to register to the draft in 1948, and I similarly found it impossible to sign a loyalty certificate required by the State University of New York when it acquired the University of Buffalo in the early 60's. In neither case was the impossibility immediately apparent to me. It became clearer through further consideration of the circumstances, of the history of the requirement, and of the meaning of my refusal. The impossibilities, that is to say, were deliberately chosen through a process of ratiocination. I suspect that the same is often true for others.

Choosing one's necessities and impossibilities is choosing one's identity. Unlike random acts or habits and quirks, what one must do and what one cannot do, and also what one is willing to do in spite of pressure or threats, play a special role in determining who one is. This is so in two ways. On the one hand, they furnish a character one can identify with and take responsibility for. This is especially true for those who follow the injunction of Socrates, "Know thyself!" and who exercise discipline in sticking to their adopted necessities, possibilities, and impossibilities. One's responsibility for who one is can then be a source of self-respect. On the other hand, one's discipline and commitments identify one as a member of the community of those who share them. Such a community may be local or geographically extended, it may be contemporary or spread through history, it may be vast or highly select, it may be intimate or impersonal, it may be actual or virtual. For most of us – certainly for me – one's various community identities are integral to one's personal identity, although one's responsibility for the community is highly variable and has entirely different dimensions from one's responsibility for oneself.

Let us return to Bush and Clinton, each "left with no choice" by a markedly weaker political power. From the perspective I have just elaborated, their claims appear exceedingly complex. I previously said that they were self-exculpatory, meant to cast historical blame for the destruction on another party. That remains one part of the picture, blaming the victim. Another part of the picture is that Bush and Clinton were delineating their characters or their political identities by deliberately choosing these necessities. In this way they were claiming credit rather than renouncing responsibility. A third aspect of their remarks is that they were attempting to build or consolidate a community, taking advantage of the tragic truth that agreement on the criteria for the use of violence has historically been one of the most powerful social bonds. In these cases, as in the more personal ones, the necessity remains contingent rather than absolute; but it is its self-consciousness, and its recognition and acknowledgement, rather than its contingency, that matters.

Contingent and adopted necessities may involve ratiocination, but at best they lead to rationalization rather than to reason. We all claim necessities in everyday life, and I see no way to avoid doing so. Some of the claims are odious, but I am disinclined to condemn *all* of them as instances of bad faith. Some of them are admirable expressions of character and/or community. Wittgenstein's point is that these necessities and impossibilities belong to our ways of responding to the world rather than to the world itself. They are not among the hard cold facts. That echo of Hume is a refreshing reminder. They are, however, part of our lives. It is integral to the examined life which Socrates recommended to bear in mind that by invoking or acknowledging necessities one is choosing one's companions and one's way of life.

Literature

al-Azm, S. 1972 *The Origin of Kant's Arguments in the Antinomies*, Oxford: Clarendon.

Diamond, C. 1991 *The Realistic Spirit: Wittgenstein, Philosophy, and the Mind*, Cambridge MA: MIT Press.

Garver, N. 1994 *This Complicated Form of Life*, Chicago: Open Court.

Garver, N. 1995 "McGuinness on the Tractatus", in J. Hintikka and K. Puhl (eds.), *The British Tradition in 20th Century Philosophy*, Vienna: Hölder-Pinchler-Tempsky, 1–15.

Garver, N. 1999 "Vagueness and Analysis", *Journal of Philosophical Research*, XXIV, 1–19

Gellner, E. 1992 *Reason and Culture : The Historic Role of Rationality and Rationalism*, Oxford: Basil Blackwell.

Ishiguro, I. 1969 "The Use and Mention of Names", in P. Winch (ed.), *Studies in the Philosophy of Wittgenstein*, London: Routledge & Kegan Paul, 20–50.

Kenny, A. 1984 *The Legacy of Wittgenstein*, Oxford: Blackwell.

Malcolm, N. 1984 *Ludwig Wittgenstein: A Memoir*, Oxford: Oxford University Press.

Malcolm, N. 1986 *Nothing Is Hidden*, Oxford: Basil Blackwell.

McGuinness, B. 1974 "The *Grundgedanke* of the *Tractatus*," in G. Vesey,(ed.), *Understanding Wittgenstein*, London: Macmillan, 49–60.

McGuinness, B. 1981 "The So-called Realism of Wittgenstein's *Tractatus*," in Irving Block, (ed.), *Perspectives on the Philosophy of Wittgenstein*, Oxford: Blackwell, 60–73.

McGuinness, B. 1988 *Wittgenstein, A Life: Young Ludwig (1889-1921)*, London: Duckworth.

O'Neill, O. 1989 *Constructions of Reason : Explorations of Kant's Practical Philosophy*, Cambridge: Cambridge University Press.

Pears, D. 1987 *The False Prison: A Study of the Development of Wittgenstein's Philosophy*, Oxford: Clarendon Press.

Stenius, E. 1960 *Wittgenstein's "Tractatus"*, Ithaca: Cornell University Press.

Wittgenstein, L. 1961 [TLP] *Tractatus Logico-philosophicus*, (tr. D. Pears and B. McGuinness), London: Routledge & Kegan Paul.

Wittgenstein, L. 1958 [PI] *Philosophical Investigations*, (tr. G. Anscombe), Oxford: Blackwell.

Causal Domains and Emergent Rationality

IVAN M. HAVEL

1. Introduction

There are perpetual attempts in philosophical and scientific literature to bypass the enigma of free will by trying to explain human action on the sole basis of physical causality, perhaps slightly seasoned with pure randomness. Obviously this affects many theorists of rationality and attracts them towards the technical interpretation of the term "rational", as it is used, e.g., in the framework of formal decision theory or, more recently, of artificial intelligence. Such an understanding of rationality has the amusing consequence that most our (human) everyday decisions turn out to be irrational, as various psychological tests have repeatedly demonstrated (Oaksford and Chater, 1998); our decisions in such a framework are rational only in cases when we painstakingly, perhaps mindlessly, execute procedures maximizing some utility function.

In contrast, in his recent book on rationality in action John Searle (2001) presents a conception of subjective, "full-blown rationality" (especially rationality in action) that presupposes, on the side of the decision maker, intentionality, consciousness, temporality, free choice, language, and selfhood. Rationality in this sense is a feature of the decision making process rather than a feature of its result. Therefore "rational" does not necessarily mean "correct", or even "reasonable" in the usual sense: if the decision is made on the basis of beliefs and desires (which then play the role of *reasons*) it is immaterial whether the beliefs are true and the desires desirable. On Searle's account, unlike the traditional philosophical views, beliefs and desires by themselves are not causally sufficient to determine rational actions, rather there is a *gap* between the "causes" of the action in the form of beliefs and desires and the "effect" in the form of the action itself. I shall call rationality in Searle's sense *intrinsic rationality* to distinguish it from *"as-if" rationality* – virtual "rationality" that people often attribute to non-human entities. In this study I present a variant of the concept of rationality, called *emergent rationality*, adopting and extending the notion of emergence as it persists over most of the last century. It is used (with somewhat vague and varying meaning) for objects, properties, or relations occurring at some more observable level of a complex system, and which are supported (some say sustained, caused, or produced) by processes and properties at some other, less visible level, but neither reducible to, nor predictable from, the less visible processes and properties (Nagel, 1961, p. 366).

While the concept of emergence is frequently used in philosophical discussions, the actual functioning of its concrete cases is usually left aside as a task for appropriate sciences. In the present study I shall try to walk the fine line between science and philosophy by proposing a certain tentative way of dealing with the traditional problem of rationality, which is

B. Brogaard, B. Smith (Hg.), *Rationality and Irrationality / Rationalität und Irrationalität*, S. 129–151.

> How can there be rational decision making in world where everything that occurs happens as a result of brute, blind, natural causal forces? (Searle, 2001, Chapter 1)

Instead of the intrinsic rationality in the sense of Searle, I will depart from the „as-if" rationality mentioned above. I believe that there are various cases of seemingly rational behavior, both in nature and in the social sphere, where the attribute "rational" cannot be so easily dismissed as nothing but a superficial anthropomorphism. The common feature of such cases is that they arise from complex multilevel systems in a non-reductive manner – which, in a sense, may also hold in the case of intrinsic rationality.

To make this clearer I shall first discuss the view of reality as fragmented into multiple "causal domains" – a partial generalization of a somewhat vague, albeit abundantly used, concept of a *level* (e.g., level of description). This makes it possible to conceive of emergent rationality as a phenomenon based on the interaction of (at least) two different causal domains, in one of which, for example, a nontrivial selection process is realized that, in the other domain, yields some sort of effectively "rational" behavior.

Such a domain-oriented approach may provide a framework for posing some questions of particular interest, such as whether, and which type of, rationality can be ascribed to collective systems of units commonly considered as *non-rational* (i.e. systems like neural networks or robot societies), and whether, and which type of, rationality can be ascribed to collective systems of *rational* individuals (systems like human societies or internet communities).

The reader who expects a solution to the problem of rationality, or at least a treatment based on precisely formulated concepts and arguments, may be somewhat disappointed. Indeed, my intention is no more than just to provoke some new thoughts and offer themes for discussion. Moreover, I could not avoid a certain sort of ambiguity in my treatment so that the reader is free to alternate between three types of reading: the psychological, the epistemological, and the ontological.

2. Four Test Examples

For the sake of easier discussions I shall first present, as a preview, four illustrative examples of particular capabilities of complex systems – complex in the sense that they inherently involve multilevel interactions. I shall later discuss the extent in which such capabilities may be viewed as cases of emergent rationality.

First example: The Clever Fluke. There is an often quoted case of the fluke *Dicrocoelium dendriticum* as an example of a parasite manipulating an intermediate host to increase its chances of ending up in its definitive host. The definitive host is a sheep, and the intermediate host is an ant. The normal life cycle of the fluke calls for the ant to be eaten by the sheep. To achieve this the fluke changes the ant's behavior by "cleverly" manipulating its neural system.[1] Whereas an uninfected ant would normally retreat into its

1. In fact, the reality is more intricate: the neural system is entered by one or two self-sacrificing fluke specimens from a group of about 50.

nest when it became cold, infected ants climb to the top of grass stems and remain immobile. Here they are vulnerable to being eaten by the fluke's definitive host (Dawkins, 1982, p. 218).

Note that the fluke species can be viewed as a quasi-material, diachronically evolving entity that is materialized in scores of individual fluke specimens. The ant-manipulating behavior is, in a sense, a property of the fluke as a species, a property with a certain distribution in a population of flukes, as well as an actual behavior of any particular fluke specimen). In the common language these distinctions are not always explicit; they become relevant for the analysis of certain emergent phenomena, as will be seen later on.

Second example: The Expressive Language. Our second example of a "clever" multilevel system is any natural language. Consider, for instance, the English language with its ability to express a variety of sophisticated ideas, such as the expression of acts not as facts, but as contingencies, as in the sentence: "There are no odds at which I would bet my life against a quarter, and if there were, I would not bet my child's life against a quarter."

We might want to differentiate between, first, particular successful speech episodes, second, the linguistic skills of the speaker, and third, the "intelligence" of the language as such. Of course, without speakers and their speech acts language would not evolve (let alone exist), and without language the speakers would be silent.

Third example: The Chess Machine. Computers have played grandmaster chess since 1980's and IBM's "Deep Blue" machine defeated Gary Kasparov, probably the best human player ever, in May 1997 match. Several times during the match, Kasparov reported signs of mind in the machine (Moravec, 1999).

This example may invoke the long-standing dispute about mentality of machines. In our context the most important distinction is between the performance of the chess playing program and the intellectual activity of people who developed the program.

Fourth example: The Rational Mind. Assume that you received a letter from organizers of a conference asking you to give a lecture on a topic of your choice. Three days later you answer affirmatively and give a title. What happened during these three days? You were probably considering various alternatives and in the course of arriving at the decision you were, I presume, quite certain, that your decision was arrived at freely, consciously, and entirely on the basis of mental reasoning. You could feel certain pressures but you would be fully aware the whole time that a different decision could have been arrived at, perhaps even one that aborted the decision making process entirely. So, I believe, this is an apt example of rational decision-making in Searle's sense.

But this is not the end of the story. You thought about it and realized that there are no sound arguments against the claim that all your

> mental phenomena are caused by neurophysiological processes in [your] brain and are themselves features of the brain,

as Searle puts it (1992, p. 1). In fact, there exists a vast amount of scientific data about processes in the brain at the neural level (and perhaps some lower levels), but there are

only speculations about how these processes might support, sustain, control, cause, or create an illusion of, mental activity.

I am now going to explore the above examples in some detail – not as to their individual nature but each as a representative of a large category of more or less similar systems. Let us first observe three general features they share and one in which they differ:

1. Each of them is a complex dynamical system that we, as observers, are used to describing and comprehending at two or more substantially distinct "levels" (phylogeny vs. ontogeny, diachrony vs. usage, programming vs. performance, mind vs. brain). The study of events at each particular level typically requires a distinctive scientific approach. I will discuss and generalize the concept of a level in Section 4.
2. The events at one level are in one way or another importantly interlinked with, or dependent on, events at another level (or at several other levels). The nature of this mutual *interlevel interaction* is not always well understood and may even require a revision of the traditional notion of causality. I will say more about this in Section 6.
3. Each of the cases exhibit a certain type of *purposive, intentional or rational behavior*, at least if we allow the "as-if" interpretations of these words. Such behavior is commonly attributed to entities (organisms, agents, persons, etc.) with respect to a single level, but as our examples suggest, it may be causally efficacious on a different level than the level that produced or sustained it. I will return to this point in Section 10.

In one interesting aspect our examples differ. Except the first example the systems involve, at a certain level, conscious beings able to act, in principle, of their own free will.[2] The difference is, however, in the actual role of the conscious and free beings and in the level on which they act. In the case of the Expressive Language the relevant activity of individual speakers occurs at the lower, "finer" level of the system while in the case of the Rational Mind consciousness is associated with the higher, "coarser" level of the system. The case of the Chess Machine differs even more: the programming activity is intentionally directed to the machine performance and hence it is not quite appropriate to talk of "levels" in this case (the same holds for other designer-artifact systems).

3. About the "As-If" Nature of Some Statements

Often people use various anthropomorphic terms, like "rationality", "purpose", "desire", "intentionality", etc., about behavior of animals, machines, social institutions and other entities that behave in an appropriately sophisticated way. In most cases such usage expresses just a tacit feeling that if we people were in the position of such entities, we would (consciously) behave in a similar way. Most philosophers cautiously indicate the metaphorical sense of the words by adding the particle "as-if" (or equivalent).

The general attitude is that the metaphorical uses of mentalistic terms are nothing but linguistic conveniences that express the beholder's external view. I propose to consider,

2. This statement is not quite substantiated because I did not made any metaphysical assumption about which entities are really conscious and free and which are not. If the reader counts flukes to be conscious, or natural-language speakers to be zombies, I have nothing against it. We would only have to look for other examples.

instead of the fact *that* a certain entity appears rational, purposeful, intentional, etc., rather *what is behind* such appearance – is there something intrinsic to the nature of the entity in question that makes it look like having such properties?

Most authors recognize only two alternatives, either the intrinsic ascription, or the "as-if" ascription, and often they are quite decided about what is what. There are few who propose further options (e.g., Haugeland, 1998). Expectedly, the main stimulus for studying the intermediate cases comes from biology, which offers unlimited number of impressive cases of apparent rational behavior.

Let us recall the statement "To achieve this the fluke specimen changes the ant's behavior ..." used in the description of our first example. Except for scrupulous scientists most people would accept the phrase "to achieve this" in this context as a legitimate and innocently metaphorical phrase used just to make the reading easier (the same holds for another, even more frequent phrase: "in order that"). We usually do not presume that such teleologically flavored expressions, when referring to (lower) animals, could be understood in their literal sense since we generally do not expect inherently intentional and conscious behavior from such a diminutive creature like the fluke.

But is this all that can be said about the phrase "to achieve this" when it is used in non-human contexts? Perhaps we might try to give it a more informed interpretation, namely, that it could possibly reveal, instead of somebody's (in our case the fluke specimen's) intention, an influence from, or at least existence of, some other, currently "invisible" level. In the particular case of the Clever Fluke it could well be the level of biological evolution. Note, however, that the teleological discourse, removed from the level of specimen behavior, may inconspicuously move to the evolutionary level where it might obtain the form of phrases like "natural selection favored this or that strategy". Isn't it so that the words "selection", "favor", "strategy" have a certain vestige of teleological meaning? Well, even here perhaps the teleological aspect can be removed by referring to the blindness of the entirely causal Darwinian variation-selection-replication process.

I was carelessly playing here with various options – in fact, with three of them: intentionality of individual animals, intentionality of evolution, and no intentionality at all – for a purpose. I wanted to demonstrate, first, that levels of description (soon to be generalized to causal domains) are relevant to the problem of rationality; second, that the attribute "as-if" should be relativized to particular levels; and third, that the nature of the sources of certain phenomena can be left open if our scope of interest is restricted.

In the following I will differentiate between *"as-if" rationality* (indicating the conviction that "nothing is behind" the phenomenon) and *apparent rationality*. I will use the latter term without, or prior to, any judgement about whether we are dealing with the case of intrinsic rationality, "as-if" rationality, or emergent rationality (to be introduced later). For example, I can ascribe apparent rationality to a concrete individual act of another person (perhaps even to myself) if the same type of act can be performed consciously (deliberately) as well as unconsciously (compulsively).

4. Causal Domains

In the preceding sections I frequently and in important contexts used the term 'level'.

Now I will generalize it in a suitable way for our purposes. In scientific as well as philosophical literature there are frequent references to concrete instances of levels, often piled up in one or another hierarchy or at least treated as if one level were "above" the other. Thus, for instance, we often read about the atomic level, molecular level, cellular level, neural level, etc., up to the psychological level, behavioral level, and societal level. However, for the most part we are left alone with our intuition about the very *concept* of a level. Judging from the usage, the only general, common feature of all levels is their epistemological meaning: they indicate much better understanding, on the side of experts, of relationships and laws *within* a particular level rather than *between* different levels.

I prefer to use the term *domain* instead of the term *level* to suppress the tacit assumption of the existence of some underlying hierarchy of levels (where some levels are usually called "upper" and some "lower"). Thus the concept of a domain is substantially more general than the concept of a level. In fact, even our everyday experiences present the world as if it were broken into various domains, each suitable for a certain type of interest and a certain type of action, and each somewhat better understood if taken separately than if combined with other domains.

In the scientific enterprise as well as in everyday life, *understanding* of the interrelationships of spatio-temporal events usually means the ability to *explain* and *predict* something. Explanation and prediction is based, in science, on *causal laws* (known or unknown), and in everyday life on the common-sense idea of one event (or state of affairs) – a *cause* – bringing about another event (or state of affairs) – an *effect* – not just by coincidence. In view of this, I shall first deal with domains based on causal relations (until Section 8 I often omit the adjective "causal").

According to the traditional view, causes are events antecedent to their effects. Moreover, some theoreticians maintain that whenever one event causes another, it does so in accordance with a general law.[3] The following introduction of the concept of a causal domain (the term was used casually by Kim, 2000, p. 69) deliberately presumes causation in the narrower sense (sometimes called the left-to-right causation). More general uses of the term "causation", e.g., the "bottom-up" or "micro-macro" causation, will be discussed in Section 6 as a type of extrinsic relation between domains.

Let us call a *causal domain* any segment (or fragment, or component) of reality within the scope of which causal relations appear to be (i.e., are presented to our knowledge as) *manifest* (obvious, apparent), *comprehensible* (intelligible), and *mutually coherent*. Or, more appropriately, they appear to be *more* manifest, comprehensible, and mutually coherent than causal relations *between different* domains are. This formulation is not meant as a rigorous definition – I am just trying to characterize the existing intuition, so the circularity should not be a hindrance. The idea of causal domains is admittedly vague[4] (depending partly on the vagueness of the underlying concept of causality), so let me clarify a little what I mean (cf. Figure 1.).

1. I do not presume, in general, that observing *concrete* instances of *causal relations* (let us call them *causal episodes*) automatically implies knowing the corresponding causal laws,
2. by saying that these relations *manifest themselves* I simply mean that in concrete situations they are easily recognizable and identifiable as causal,
3. by saying that they are *comprehensible* I want to imply that most of us accept them without a need for further explanations, or without surprise,
4. *mutual coherence* of the relations means that they belong to one common, apparently consistent, *causal network*. In particular, various causal episodes may fit together and form arbitrarily long *causal chains*.

3. In this paper, I am not concerned with the issues of the nomological character of causality.
4. The vagueness of the idea may not impede the fertility of some of its applications.

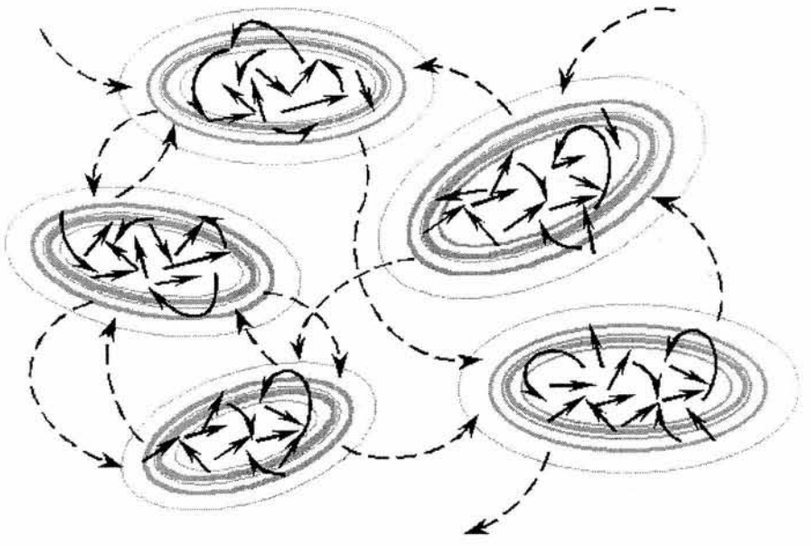


Figure 1. Causal domains. Causal relations within a domain are more manifest, comprehensible and mutually coherent than between different domains.

The concept of a causal domain can be used with respect to any individual (private) knowing of the world, as well as to the collective, scientific knowledge. Examples will make it clearer.

On the individual side, we have our everyday experience of temporarily restricting our attention to the (presumed) causal context of an activity that we are engaged in or of a problem we want to solve. Let us say, for example, that you have a problem with the engine of your car. In order to fix it you must take into account the various causes and effects relevant to the functioning of the engine. But you would probably not take into account, say, the causes and effects relevant to the chances of it raining the next day. Later, though, when you cut your lawn, it may be the other way around. Note that such "causal domains" are more or less shaped by your individual knowledge.

More interesting examples of causal domains are the particular subject areas of natural science (like, for instance, quantum physics, molecular biology, or evolutionary biol-

ogy) and the fields delimited by different research methods. More or less anything we find in scientific and philosophical discourse designated or recognized as a "level" falls under the concept of causal domain. In scientific contexts causal domains and levels of description are typically shaped by collective knowledge.

Each causal domain may be viewed as a world unto itself: it has specific individuals, universals, properties, aspects, relations, laws, etc. Such things may be peculiar to one particular domain – then I shall call them *endemic* to the domain. Other things may belong to, or be meaningful in, several, perhaps many, domains (so-called *multidomain* entities and *shared* concepts).

A few examples of the latter case may be useful. First, consider a natural physical phenomenon like lightning. Whether as a token (concrete event) or as a type (the generic concept), it can be conceived as a multidomain entity. It, so to say, "penetrates" through many domains differing in scale: from the macroscopic scale, as we see it in the sky, to the microscopic scale of its physical description as a collective flow of charged particles. The difference in scales is more than twelve orders of magnitude which makes it difficult (for an observer) to comprehend this phenomenon as one single entity. Another, more interesting example is the mental state of, say, fright. It may be studied in the behavioral domain (as a pattern of behavior of most animals), in the endocrine domain (as a release of adrenalin), in the mental domain (not quite pleasant first-person experience), in the genetic domain (as a trait carried on certain genes), in the evolutionary domain (as a factor in natural selection), etc.

Such multidomain entities may be contrasted with cases of generic concepts and properties *shared* by various domains, i.e. having analogous, or even the "same", meaning in various domains, partly due to their generality, partly due to the limitations of our language. Obvious examples are the concepts of space, time, causality, and most mathematical abstractions. Many domains, especially those that are a by-product of science, are not directly accessible to our intuition (think, e.g., of the domain of elementary particles). We can describe them only mathematically or by metaphorical transfer of vocabulary, borrowed from perceptionally accessible domains (e.g., the "spin" of elementary particles or the "collapse" of a quantum wave function).

A typical causal domain lacks sharp borders, and what we count or do not count to it depends on how far we are able to extend the connected network of mutually coherent causal relations. On the individual side it is often related to how far our "sight" is able to reach. For a particular scientific discipline this "sight" is not so much dependent on the visual field and viewpoint of a concrete observer but rather on concepts, quantities, laws, and paradigms that are peculiar or significant for the current state of knowledge in this discipline. Naturally, such a discipline is limited, but at the same time it lacks a clear border. So, instead of a border, we deal with the concept of a *domain horizon*.

This brings us to the challenging question: are causal domains really "out there" in the world or not? I think both, in a sense. There is something inherent in the fabric of reality that makes it easier for us to cope with the world in relatively separated regions (levels of description, areas of interest, spheres of knowledge). At the same time there is something inherent in human nature (limited scope of conscious attention, physiological limitations of perception, etc.) that forces us to approach the world in fragmented way. Thus, to the same extent that we are realists about the difference between trees and forests, water drops and clouds, bees and beehives, neurons and brains, and causes and their effects, we should be realists about causal domains. Yet, undoubtedly, we have a great amount of freedom to fix the details of such decomposition. Our picture of the world is a dynamical outcome of a never-ending circular hermeneutic process: our world is *enacted* (Varela *et al.*, 1991).

Let us imagine that a certain collection of causal domains can be ordered into a linear sequence with the help of some natural or artificial ordering characteristic. For example, we may sort the domains according to the dominant spatial and/or temporal scale of typical objects or processes in each domain (Havel, 1996), according to the structural (mereological) subordination or functional dependence of entities belonging to different domains (Scott, 1995), or according to any other suitable characteristic of the domains in question. Only when the collection of domains is so ordered, is it appropriate to talk about a *hierarchy of levels*.[5] The term *level*, used for any of the domains in the hierarchy, is therefore relative to the chosen ordering characteristic. In our account the term "level *of description*" is derivative and the difference between *higher* levels and *lower* levels is then implied. So, for example, in biology we can talk about a large hierarchy of domains ranging from the molecular level, through the level of individual organisms, up to the level of the biosphere. Or, in the time-scale hierarchy, we can separate the phylogenetic domain, the ontogenetic domain, and the domain of behavioral episodes. On the other hand, the manner of treatment of the mind as a "higher level" and the brain as a "lower level" in a certain hierarchy is, from the point of view explained above, somewhat dubious: there is no obvious ordering characteristic applicable there (in fact, this is one of the reasons for my preference of the concept of a causal domain).

We are usually able to shift our focus from one domain to another: in daily life we do it all the time; in science it may depend on profession: scientists in one field just don't feel quite "at home" in other fields. At the same time it is difficult, if not impossible, to keep several domains in sight simultaneously when they are sufficiently "distant" from each other, while "neighboring" causal domains cannot be sharply separated.

5. The Mental Domain

We are used to ascribing causal nature not only to physical relations (in the broad sense of the word "physical") but also to relations involving our mental states. It is therefore appropriate also to take into account *mental causal domains*. For example, suppose you become frightened after seeing a snake and that makes you freeze up. You might sense this episode as a causal chain and describe it in a causal language, as if your perception directly caused your fright and your fright directly caused your immobility. No allusion to the physiology of your body is required.

The subjective mental domain (or "inner world") of a person includes experienced mental states and their causal relations (when something causes a mental state, or when a mental state causes something else). Hence, it also includes the mental representation of

5. To simplify matters we do not consider here abstract "hierarchies" formed by mutual inclusion of domains.

that part of the physical world that can be affected by intentional acts, and also affect mental states, of a person. In particular, it includes the person's body and bodily movements.

We can distinguish (with, e.g., Chalmers, 1996) two kinds of mental domains, the *phenomenal domain* (of first-person experience accessible to consciousness) and the *psychological domain* (the causal or explanatory basis for behavior described in the third-person way). Within the phenomenal domain we can further differentiate the perpetually changing subdomain of consciously attended mental states from the realm of the unconscious. The assumption that conscious beings have phenomenal mental domains that differ from each other only in their concrete contents (the experienced instances of mental states and experienced instances of causal relations) allow us to talk "objectively", or more appropriately, "intersubjectively", about the mental domains of other persons.

An instance of a causal relation in the phenomenal causal domain is undeniably "true" *within* that domain in a similar sense as a hallucinated object is undeniably "seen" by the hallucinating subject. Consequently, it may not be compatible with causal relations in the psychological or other domains.

6. Connections Between Causal Domains

So far I have been somewhat reticent about the ways events in one domain could influence events in another domain – except for the vague consequence of the "definition" (of causal domains), according to which if the influence between domains is causal then it is, in general, less manifest, less comprehensible and less coherent than causal influences *within* each domain. This does not mean, however, that the interdomain relationships should be considered weak or irrelevant. In fact, one of the motivations behind our approach is to allow for certain kinds of efficacy between domains that may perhaps have a different nature than the classical causal relations (and laws).

Let us see what present-day science has to offer. In the past few decades, scientists have dealt with various situations that characteristically involve two or more different levels (or domains). Besides already mentioned statistical physics, there are theories of structural and/or shape interaction (e.g., of large molecules), quantum effects (e.g., nonlocal quantum phenomena), and various cooperative, non-linear, chaotic, synergetic and emergent phenomena. For some types of influence we lack (at the current state of knowledge) a formal description or even an intuitive grasp. The most peculiar case is, of course, the psychophysical (mind-body) interaction.

First, let us make some terminological distinctions. I will use, in general, the term *efficacy* for any conceivable influence or dependence, without any apriori claim about its physical (or nonphysical) nature and even about its direction. Thus a connection between different domains is efficacious if, in virtue of it, events in any one domain can bring about events, or affect the actual state of affairs, in the other domain. Hence, (interdomain) *causality* is a special case of (interdomain) efficacy. In the case of causality we automatically assume an explicit direction of effect; whenever I also want to include reciprocal or mutual efficacy (whether causal or not) I prefer to use the more general term

interaction. Kim (1974/1993) analyses various cases of what he calls "noncausal connections" as being dependent on the structure of events or states of affairs in question.

Among *non-efficacious* connections between domains we may list some *logical* or *analytical relations* that often depend on the way we understand the nature of "events", "state of affairs", "properties", etc. The important concept of *supervenience* is defined by some authors essentially as a logical relation (Kim, 1974/1993), while some other authors treat it as just a feature of causality (Searle, 1992, pp. 124–126). The term *correlation* and *parallelism* may be considered to be tentative words for phenomena currently lacking a causal explanation.

Across this classification are *systemic* connections based on the existence of a complex multidomain system (cf. Section 9). For instance, the important phenomenon of *emergence* can sometimes be treated as a directed causal relation and sometimes as a non-efficacious and timeless systemic relation (the type of the treatment depends on our point of view).

Due to the nonexistence of a precise concept of an extension of a domain there is no exact dividing line between *endemic* causality and *interdomain* causality. The endemic causal episodes often have temporal character (the cause precedes its effect; hence the "left-to-right" intuition for this kind of causality). Interdomain causation typically involves at least two different domains that are often viewed as different levels in a certain scalar hierarchy. The typical cases of interdomain or interlevel causality are *upward* causation and *downward* causation.

When we deal with an efficacious relation between different domains (or, more commonly said, between different levels), we have much to learn about its nature before claiming that it is causal, or even that it is an instantiation of some general causal law. We should not take it for granted that ideas used in thermodynamics for explaining heat or in evolutionary theory for explaining mimicry can be applied everywhere, including the theory of mind.

7. Biological Naturalism and Causal Gaps

John Searle (1983) formulated his *biological naturalism* as the thesis that "mental states are both *caused by* the operations of the brain and *realized in* the structure of the brain." This leads to the interpretation of causal sequences at different levels as "not independent causal sequences, but the same causal sequence described at different levels" (Searle, 2001, Chapter 9).

In the framework of causal domains we cannot say so easily "*the same* causal sequence", at least in cases without a clear theory of interdomain connections. Perhaps we could be more cautious and render the situations that Searle had on mind as cases of "causal parallelism": certain causal episodes in one domain X appear to be accompanied by certain causal episodes in another domain Y.

Consider, for instance, that your percept of a snake caused your fright (causal relation in the mental domain). In parallel, as if by coincidence, a certain pattern of neuron firing in your visual cortex caused some other pattern of neuron firing in your motor cortex, inhibiting your muscular movement (an instance of a causal relation in the neural domain).

We don't know yet what the *nature* of the seemingly efficacious interdomain connection is, nor even its direction (or ever if it has a direction at all), yet the *existence* of a connection is a sound scientific hypothesis. Talking about causal parallelism in such a case is scientifically vague but intellectually helpful.

In fact, it leads to a difficult issue of making sense of the possibility of *free* decision making in the mental domain. Following up our snake example, consider that when seeing the snake you overcome the instinct and launch a conscious deliberative process aimed at a decision about your future behavior, for instance whether to stop moving or run away (there may be good *reasons* for both options). Is the idea of causal parallelism still applicable?

John Searle (2001) elaborated a theory of intrinsic rationality in action presupposing, on the side of the decision-maker, conscious awareness of the existence of alternatives for a free choice. Thus, instead of being causally determined by an antecedent set of beliefs and desires, rational decision making presupposes a *gap*. In Searle's words, it presupposes

> [...] a gap between the set of intentional states on the basis of which I make the decision, and the actual making of the decision. That is, unless I presuppose that there is a gap, I cannot get started with the process of rational decision making. [...] We presuppose that there is a gap between the "causes" of the action in the form of beliefs and desires and the "effect" in the form of the action. This gap has a traditional name. It is called "the freedom of the will" (Searle, 2001, Chapter 1).

Searle's account of the human rational process is primarily presented within the framework of the subjective mental domain of a person. In that domain, the gap can be viewed as an opening, or play,[6] for some sort of external intervention – the free will (or what is subjectively sensed as the freedom of the will). The actual "sources" of free will are beyond the horizon of the mental domain, and thus the options are open whether to take it as a primitive principle or to believe in some natural explanation, perhaps exposed to scientific investigation. In Figure 2 there is a schematic "parallelogram" illustrating the situation (cf. Searle, 2001, Chapter 9.)

In the scheme presented in our framework the phrase "external intervention" may actually mean "intervention from another causal domain." This, incidentally, may help to sort out various physicalistic and reductionistic theories. What fills the gap? Searle's answer, "Nothing", can be interpreted as: "Nothing in the mental causal domain." This opens the question of the nature of interactions with other domains, which either may or may not fill the gaps deterministically.[7]

I am not going to entertain this issue here, and I mention it just to motivate the idea of extending the concept of a gap to arbitrary causal domains. Let us define, rather vaguely, a *gap in a causal domain* as any opening in the network of causal relations (in that domain) for efficacious influences from another domain(s). Schematic picture of a gap is in Figure 3.

6. Let me note that the Czech word 'vůle' means both 'will' and 'play' (= clearance).
7. In the last chapter of his book (2001), Searle proposes two possibilities of explanation of human rational behavior (psychological indeterminism either coexists with, or is matched by, neurobiological indeterminism).

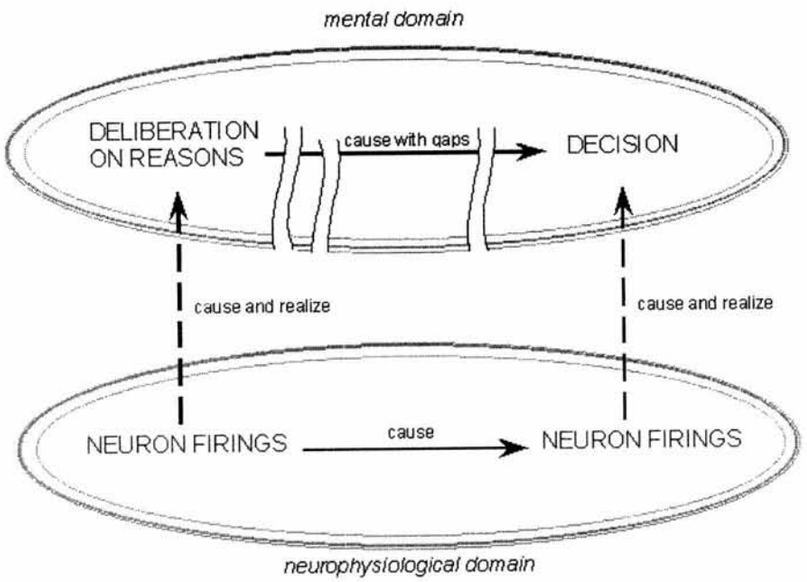


Figure 2. Searle's parallelogram with experienced gaps. (According to the compatibilist hypothesis gaps appear only in the mental domain, neurophysiological domain is deterministic.)

For example, in the case of the fluke we may ask what is the actual origin of the fluke's solution to the nontrivial problem of reaching the sheep's digestive organs. This question may lead us to the evolutionary domain, but there we can find a "gap" in the linear flow of causally connected events: a "play" where Nature (or Evolution) could make a "choice" among a practically infinite number of alternative solutions to the fluke's problem. The gap has been bridged by blind chance (this is one point of view) or by rational design (another point of view).

From one's "view" within a domain, gaps appear, so to say, on the horizon of the domain (cf. Section 4). Under a suitable extension of the domain, it is possible that some of the gaps would cease to exist while others might emerge. Thus the concept of a gap is always domain-relative.

For explanatory purposes within a particular domain, such gaps are usually "filled" with the help of various default assumptions. In science it is often the assumption of randomness. For example, biologists, when they discuss neo-Darwinian theory in the evolutionary discourse, would make the assumption of random occurrences of mutations (read: the gaps in evolutionary domain are filled with random events). This sidesteps looking

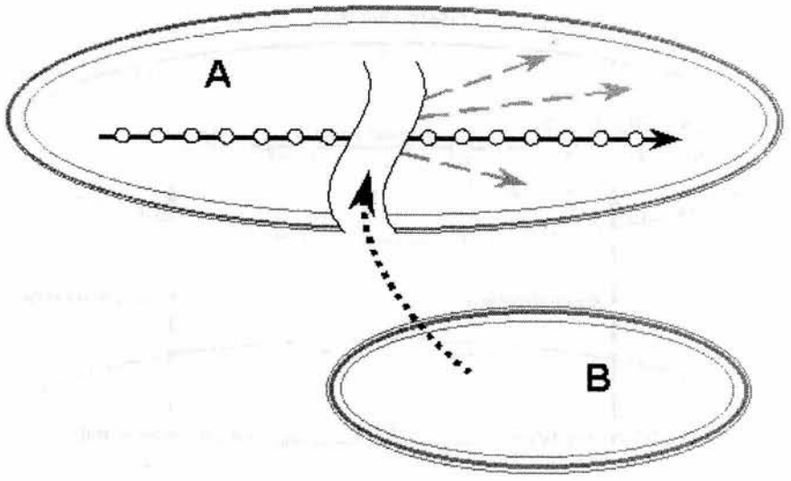


Figure 3. Causal domain with a gap: an opening in the network of causal relations in domain A for influences from domain B.

for another causal domain in which each particular mutation could be explained as the outcome of a certain causal chain. Similarly we may read Searle's theory of gaps as a way of saying that the default assumption for the gap filler in the mental domain requires the concept of the self.

8. Causal Domains, Explanatory Domains, and Rational Reasoning

Let us refer to the following observation (by Kim, 1974/1993):

> It is congenial to the broadly realist view of the world that most of us accept to think of the network of causal relations in the world as underlying, and supporting, the network of explanatory and other epistemic relations represented in our knowledge of it.

In particular, we expect many "why" questions about events in the world to be answered by pointing, at least indirectly, to their causes (rather than to their effects). So, for example, when asking: "Why do pebbles have a smooth surface?" we will be satisfied with the reply, "Because frequent collisions with other pebbles stripped them of projections on the surface." We might be less satisfied with the reply: "Because a smooth surface helps them to persist in the stream," or with the trifle: "Because they are defined so" (even though these answers are true as well).

Now let us recall the main aspects of causal domains, namely, that causal relations *within* domains are manifest, comprehensible, and mutually coherent more than causal relations *between* domains are. If our knowledge of the world is structured into such domains, then the "network of explanatory and other epistemic relations" is also structured into such domains. Thus we can introduce the concept of *explanatory domains* within which the explanation of concrete events and facts is easier, more direct, or more acceptable than explanations between explanatory domains. In the first approximation, if we consider only the ordinary, left-to-right type of causality and the corresponding (let us say "right-to-left") explanations, there is a relative match between causal domains and explanatory domains. This match may be understood as a result of a continuing mutual interaction of our ontological views with our epistemic conceptions.

There are, of course, other kinds of explanatory relations. Some of them may refer to various types of interdomain connections; their explanatory force (subjectively valued) obviously depends on our acceptance or nonacceptance of the type of interdomain connection referred to. Other, particularly interesting explanatory relations may refer to voluntary (human) behavior and accept explanations that cite *reasons* rather than just causes.

Evidently, the explanatory success of an answer to a "why" question concerning a voluntary action may crucially depend on the characteristics of the inquirer's explanatory domains, and the veridicality of the answer may depend on the agent's choice of a particular domain as a background for reasoning about the action. Indeed, talking about reasons is meaningful only if they are somebody's reasons, and that the "somebody" has to be a free entity with intentions (a *self* in the Searle's sense or an *existence* in the Heidegger's sense). Only then, the argument would go, can we expect the ability to weigh various pros and cons of alternative choices from such an entity.

To develop a notion of explanatory domains that would include domains of rationality (i.e. causal domains in which, besides causes, reasons are also manifest, comprehensible, and mutually coherent) appears to be a nontrivial assignment. Fortunately, our task is somewhat easier since the types of *apparent rationality* we are concerned with allow us to assume the intersubjective position.

9. Multidomain Entities and Systems

I have already mentioned that there are objects and events that can be treated as multidomain entities. Our previous examples (of lightning and fright, cf. Section 4) only demonstrated the relevance of various domains for the description of such entities. There are, however, cases for which various domains play an essential role from the functional point of view. The term *multidomain system* (and its particular case, the *hierarchical system*) might therefore be more appropriate, whether it is used in the ontological sense or in the epistemological sense. In fact, it is often the case that the collection of epistemologically relevant domains of such a system (e.g. the hierarchy of levels of description) more or less coincides with the collection of ontologically relevant domains (the hierarchy of organizational and functional levels).

Particularly interesting for us are multidomain systems that are complex, so to say, twice over: first, they are complex with respect to each particular domain relevant for the system, and second, they are complex due to the presence of a web of multifarious efficacious interactions between these domains. Not just the presence, but the durability of the

whole system may crucially depend on these interactions. The systems and entities to be discussed here typically evolve over time in various ways (depending on which domain is considered). For simplicity, however, I will not pay particular attention to their origins.

Obvious examples of complex multidomain systems are living organisms, ecosystems, social organizations, or computer systems. In this section I reconsider the examples from Section and present each as a multidomain system emphasizing the types of interdomain interactions that are for one reason or another interesting.

First example: The Clever Fluke (revisited). In the case of the fluke, two particular domains of biological discourse are clearly involved: the *evolutionary domain*, in which the main objects of study are animal species (and their evolution), with underlying time scales of millions of years and more, and the *domain of individual animals*, with underlying time scales from seconds to years. For some purposes another domain, the *domain of populations*, can be distinguished (with mostly statistical characteristics). The domain of individual animals can be further divided to the *ontogenetic* (or *developmental) domain* of individual life (months and years) and the *episodic behavioral domain* (seconds and days). One could (I will not do it) further add various physiological and biomolecular domains supporting the behavior.

The "distance" between the evolutionary domain and the domain of individual animals is so huge (especially due to different time scales) that changing the discourse from one to the other requires a radical mental shift. When we pay attention to one of the domains, the other almost disappears from our sight; consequently, we are not confused when the same words are used in both domains. So, for example, the statement "On islands smaller animals grow, while larger animals shrink" is logically meaningful in both domains (while its factual meaning is rather different).

Let us consider, theoretically, a conglomerate of domains relevant to the fluke strategic behavior, from the evolutionary domain of the fluke species to the episodic behavioral domains of every fluke specimen, all that combined into one complex multidomain system. Let us call it "the Fluke System". Assuming, for instance, the ordinary Darwinian theory[8] we can easily identify the nature of mutual interactions between the two salient domains of the Fluke System. It is the evolutionary domain of the fluke species on the one side, and the domain of behavioral and life episodes of individual fluke specimens on the one side. The interactions can be roughly described as follows: the inherited properties shape individual behavioral patterns, and conversely, the successes and failures of behavioral episodes have a cumulative effect on the hereditary properties of the species (a more precise description would take into account the domain of individual life stories and the domain of populations).

Now we can claim that the ant-manipulating strategy is an essential property of the *whole* Fluke System.

Second example: The Expressive Language (revisited). Let us consider our second example of a natural language; call it "L". A completely different time scale applies (1) to L

8. Here it is irrelevant whether the Darwinian theory yields a correct account of the fluke behavior or not.

as a historically evolving diachronic entity, (2) to the learning process of speakers of L, and (3) to any particular act of uttering (or writing, listening or reading) a sentence in L. There is, nevertheless, a two-way interaction between the corresponding levels: for instance, each concrete utterance chooses words in L, obeys grammatical rules of L, and follows the habits prevailing at the respective historical moment among the speakers of L. On the other hand, the vocabulary, grammar, and habits of L evolve over long periods of time, and are subject to the accumulated influence of many actual utterances. Because of these interrelationships, we should realize that these different components are just different facets or manifestations of a single multidomain system. Let us call it "the Language System".

Similarly, as in the case of the fluke, the intelligent "inventions" of language, like the subjunctive mode in English, is not just a property of concrete speech episodes, nor of individual speakers, nor of the nonmaterial historical entity L, but rather a property of the Language System as a whole. The Language System interestingly differs from the Fluke System. Some of the domains of the former involve conscious intentional entities, namely the minds of the users who *intentionally* use the various features of L, like the subjunctive mode, and in this way *unintentionally* helps to preservation it in the language (I will return to it in Section 10.)

Third example: the Chess Machine (revisited). Our first idea may be to view a computer as a multilevel system with functional organization of the hierarchy of levels. For instance, classical computers comprised several clearly distinguishable hardware levels (from electronics to central processing units) as well as software levels (from machine code to programs in a high-level language). The conceivable complexity of such a system led some thinkers to an undue optimism about the possibility of the spontaneous emergence of mental phenomena in a hierarchically organized machine (cf. Hofstadter, 1979).

We should take into account, however, that contrary to our previous examples, the hierarchical organization of the computer is artificially constructed all the way down to the "silicon" level and also the types of interlevel connections are part of the prior design project. Thus, the designer(s) should be taken into account.

Let us consider "the Chess-Machine System" including tentatively the following most relevant causal domains: (1) the *physical domain* of the execution of the chess-playing program; (2) the program *performance domain* (described in the language of chess); and (3) the *programming domain* subsuming the relevant part of the mental domain of the programmer[9] while working on the chess-playing program. The first two domains are concerned with particular chess matches (actual or potential) whereas the programming domain involves the whole process of the development of the program. Thus, all concrete decisions made by the programmer (for instance, which particular utility function should be implemented) are part of the system as much as the physical processes in the computer.

What are the interdomain connections in this case? Here the situation is somewhat complicated by the involvement of the programmer's intentionality. In fact, it is the source of explicit intentional links from the programming domain to the performance domain. Moreover, it also establishes mutual connections between the physical and the per-

9. Actually of a team of programmers. I am using the singular just for simplicity.

formance domains, whereby certain objects in the former are assigned appropriate meanings in the latter. The linkage from the performance domain to the programming domain is obvious: the outcomes of actual or imagined chess matches may cause further development of the program.

So far so good, but *who* actually defeated Kasparov? It seems that analogously to the previous cases we should not isolate one of the domains and look for a "responsible" entity in there. Neither the computer alone defeated Kasparov (as a machine it could not do more than just to follow the laws of physics), nor the programmer (nobody expects him to be an excellent chess player). What remains is the Chess-Machine System as a whole.

Fourth example: The Rational Mind (revisited). The ant-manipulating strategy of the fluke, the subjunctive mode of a language and the winning chess program were just three concrete examples of phenomena in multidomain systems with rather surprising degree of sophistication. This may motivate us to wonder about the "system" that is sophisticated *par excellence*: "the brain and its mind".

Perhaps we would like to start (as many theoreticians do) with the brain and identify an appropriate hierarchy of levels in which each level could have its characteristic language of description, type of described phenomena, and its own causal relations (hopefully even causal laws). If the brain could be compared to the classical computer (as some believe), it would be easier: even if it is rather difficult within one view to embrace all functional levels in the computer, we are at least able discern these levels and specify the way they interact. The human mind affords us, however, a different story. Even if neuroscientists can describe very thoroughly some of the causal domains of the (human or animal) brain, these "known" domains are separated from each other by a large hiatus in knowledge. Thus, it is more our wish than a real possibility to imagine the functional organization of the brain in the form of an intelligible hierarchy of levels. Even more audacious is the wish of some materialists and emergentists that our genuine mind would occupy a certain sufficiently "high" level in the very same hierarchy.

I take both wishes with doubts. Even if we were able to identify organizationally and functionally "dense" (fitting with each other) collection of causal domains, it would not guarantee that the web of interdomain interactions would allow us to arrange them into a simple hierarchy. And even if such hierarchy existed, this would not imply that the mental level belonged to it.

What differentiate living organisms (and brains) from machines are the very existence, interplay and mutual interaction of a large number of different causal domains. Thus, we have again a multidomain entity, let us call it "the Brain-and-Mind System", that seems to comprise many different causal domains. The complexity of the system is not as much related to the number of domains as to the fact that they are densely "packed" within relatively few spatio-temporal scales (even if some of the domains involve highly parallel functioning of an exceedingly large number of active units). Both density and interaction are crucial features here. The density makes it difficult to study the domains individually and the interactions between domains may require fundamentally new scientific approaches.

Analogously as in the previous examples, we may view mental phenomena as though they are sustained by certain emergent properties of the whole Brain-and-Mind System.

For this, however, we need to develop a concept of emergence more general than the ordinary one.

10. Second-Order Emergence and Rationality

The examples in the preceding chapter implied that most impressive cases of apparently rational behavior arise in complex multidomain systems – they deserve to be called "complex" in the sense that their existence and durability requires nontrivial interactions between various causal domains (levels). Usually such systems are not conceived as one single entity, perhaps due to the fact that the relevant domains (levels) are conceptually isolated and/or that they substantially differ in their characteristic spatio-temporal scales.

An exemplary situation is sketched in a schematic way in Figure 4. Roughly speaking the scheme illustrates the history of a kind of entity in domain A (to be called "upper level") in mutual interaction with a population of individual instances (episodes, occurrences, tokens) of the same kind in domain B (to be called "lower level"). Let us ignore possible direct interactions between individuals in domain B. It is conceivable that under a change of perspective the same scheme repeats downward, upward, or sideward forming a larger multidomain system.

In reference to our examples, the scheme in Figure 4 may correspond to two main domains of the Fluke System or two main domains of the Language System.

For the attempt to look for a possible background of apparent rationality the usual concept of emergence (as a phenomenon at one level, supported or produced by events at another level) turns out to be insufficient. Elsewhere I proposed a new conception of emergence, the *second-order emergence* (Havel, 1993). Roughly speaking, an entity is second-order emergent if it arises from global interaction or "cooperation" of many domains of a complex multidomain system.

I am suggesting here that the term *emergent*, when applied to rationality, (or to purposiveness, inentionality, etc.), should imply the second-order emergence. This does not rule out attributing emergent rationality to endemic entities in a particular causal domain (that is, in a domain-specific discourse). However, then it expresses (unlike the "as-if" ascription, which suggests that there is "nothing behind" an appearance) our understanding that *behind* the appearance there is something more: a multidomain system with nontrivial interdomain interactions.[10] Thus it is, in effect, much stronger attribution of rationality than a mere "as-if" attribution. On the other hand, emergent rationality may be a weaker phenomenon than intrinsic rationality (in Searle's sense) because no claim is made about the necessity of involvement of a conscious self associated with the domain in question.

We may demonstrate our conception of emergent rationality with the Fluke System, for simplicity reduced to two interacting levels as in Figure 4. As observers, we would admire the apparently rational ant-manipulating behavior at the specimen level (domain B). But there it is (according to current views) just a fixed genetically preprogrammed behav-

10. An alternative formulation may associate emergent rationality solely with the whole multidomain system and in the domain-specific discourse only use the term figuratively.

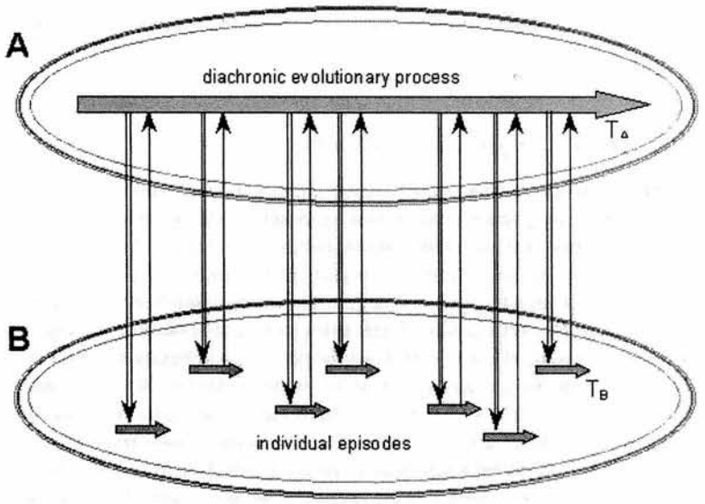


Figure 4. Two-level system S_{AB} evolving in domain A (in time scale T_A) with individual episodic events in domain B (events in typical time scale T_B). The upper-level process sets the parameters for episodic events (double arrows); the events contribute to the evolution of the process (single arrows).

ior. At the same time, we know that the program is a result of much slower upper-level evolutionary process (in domain A). This process, in turn, could have only made the "decision" about the proper ant-manipulating strategy after a long history of testing various alternative strategies with individuals in domain B. Thus the apparent rationality is the outcome of "cooperation" of both domains significantly separated in time scales.

The example hints at the possibility of the decision being made in one domain while the reasons justifying it belong to quite a different domain. This possibility opens an interesting area of investigation: *How to understand a multidomain system that exhibits apparent rationality in one of its domains while it involves conscious and intentional agencies in quite another domain?*

One category of such systems, exemplified by the Language System, includes various social institutions and organizations: legal systems, political structures, corporations, science, art, games, etc. They have the common property that their overall history (in a larger time scale) depends in a certain known or unknown way on a multitude of episodic events or acts controlled by conscious agencies at a "lower" level (in a smaller time scale). These agencies have specific intentions and goals and enjoy their freedom of choice.

Consider again a two-level system in Figure 4. Assume that the episodic events are at the lower-level (domain B) and that the upper-level evolving process (in domain A) is neither random nor under anybody's direct control. Rather it is dependent on the cumulative effect of the lower-level episodic actions made by intentional agencies. In a special case, when these agencies are unaware of the dynamics at the upper level, the whole system has similar non-personal character as many other systems encountered in nature. This may be the case of the Language System, if we assume that the speakers, who intentionally use various features of the language, are unaware of possible reverse effects of their speech acts on the language as such.

Now assume that the agencies at the lower level, besides being consciously concerned with individual episodic events, also know about the existence and dynamics of the upper-level global process. Then several cases can be discerned. First case: the lower-level agencies have no desires to influence the upper-level process (they "think and act locally"); second case: they have such desires but they have no idea how to realize them; third case: they believe they know how to realize such desires and they behave accordingly ("think globally, act locally"). Of course, they may be mistaken in their beliefs – which happens to be a frequent case in human societies.

In the last case, it is conceivable that individual decisions at the lower level collectively (and successfully) favor some development at the upper level (on which the agencies may agree, perhaps only by a majority). We might want to talk in such a case about *emergent collective rationality*. Even if it is not based on any upper-level consciousness, in the minds of the lower-level agencies (now in the role of observers) the upper-level dynamics may well produce a (false) impression that it is under control of a virtual upper-level rational agency. The impression may be even supported by the fact that the influence of each lower-level individual act on the behavior of the upper-level process may be infinitesimal due to the enormous number of other lower-level individual acts that jointly participate in affecting the direction of the overall process.

Let us turn next to another category of multidomain systems exemplified by our second two examples, the Chess-Machine System and the Brain-and-Mind System.

For the first case, consider again the question "*Who* defeated Kasparov?" in reference to the concrete May 1997 match (to avoid discussion of the game of chess in the large, or the field of AI in the large: this would bring us back to the previous category).

In this case it is popular to attribute the achievement – and, in general, the competence of rational decision making – either to the "Deep Blue" machine alone, or to the programmer (programming team) alone. However, in the framework suggested in this study, we should consider the whole multidomain system it its entirety, including the physical domain, the performance domain and the programming domain (perhaps, some other decompositions may be appropriate). The apparent rational behavior in the performance domain can be then grasped as a second-order emergent property of the whole Chess-Machine System.

The issue of machine "mentality" has been extensively studied since the ascent of programmed computers an it is not my aim here to sort and review various diverse views. It seems natural and compatible with our approach to talk in this case about *derivative rationality* in a similar sense as Haugeland (1998) talks about derivative intentionality. The person who creates the program is well aware of the ends of his actions (the programming domain includes the relevant part of his mental domain) whereby these ends are formulated in the language of the performance domain (e.g., what amounts to be a winning strategy). Thus, the apparent rationality in performance of the machine is derived from the rationality of the programmer.

Finally, let me shortly mention the Brain-and-Mind case that is, indeed, the most in-

teresting one, even if it is not, in fact, quite in the scope of this study in which we mostly deal with other than intrinsic rationality. What can make it attractive for us is that with the mental domain both the intrinsic rationality (involving consciousness) and the apparent rationality are associated, while the brain domain (or a variety of biological-physical domains related to the functioning of the brain) is just a structure composed of apparently non-rational elements (neurons and their collectives).

In view of the cases mentioned earlier, we may pose a question to what extent the concept of emergent rationality, based on the idea of second-order emergence phenomena in complex multidomain systems, may help us to understand also the Brain-and-Mind System. If we believed, like earlier Searle, that our (individual, human) mental phenomena are caused and sustained by "blind" neurophysiological processes in the structure of the brain, why should we be so reluctant to ascribe analogous phenomena, intrinsic rationality included, to some higher structures (languages, human organizations, etc.) based on, or composed of, an immense number of mutually interacting (even possibly intentional and rational) lower-level individuals?

Or conversely: assume that neurons (or other basic units) in the brain are all conscious, intentional and rational little creatures. What difference would it make if these creatures, in addition to their local interests, were aware of the existence of upper-level phenomena and deliberately influenced them? (Let me parenthetically remark that such a question may become rather relevant in the future Internet society.)*

Literature

Chalmers, D. J. 1996 *The Conscious Mind: In Search of a Fundamental Theory*, Oxford: Oxford University Press.

Dawkins, R. 1982 *The Extended Phenotype*, Oxford: W. H. Freeman, 1982.

Haugeland, J. 1998 *Having Thought: Essays in the Metaphysics of Mind*, Cambridge, Mass.: Harvard University Press.

Havel, I. M. 1993 "Artificial Thought and Emergent Mind", in *Proceedings of the International Joint Conference on Artificial Intelligence '93*. Denver, CO: Morgan Kaufman Professional Book Center, 758–766.

Havel, I. M. 1996 "Scale Dimensions in Nature", *Int. Journal of General Systems*, 24, 295–324.

Hofstadter, D. R. 1979 *Gödel, Escher, Bach: An Eternal Golden Braid*, New York: Basic Books.

Kim, J. 1974/1993 "Noncausal connections", *Noûs*, 8, 41-52, also in: Kim, J. 1993 *Supervenience and Mind. Selected philosophical essays*, Cambridge: Cambridge University Press, 22–32.

Kim, J. 2000 *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*, Cambridge, Mass.: The MIT Press.

* The research is partly supported by the institutional research project MSM 110000001 of the Czech Ministry of Education. I owe a special debt to my colleagues in CTS for interesting conversations about the subject. Thanks are also due to Kevin Coffey for correcting my English in sections that the reader may easily recognize.

Moravec, H. 1999 *Robot: Mere Machine to Transcendent Mind*, Oxford: Oxford University Press.

Nagel, E. 1961 *The Structure of Science: Problems in the Logic of Scientific Explanation*, New York: Harcourt, Brace & World.

Oaksford, M. and Chater, N. 1998 *Rationality in an Uncertain World* (Essays on the Cognitive Science of Human Reasoning), Psychology Press.

Scott, A. 1995 *Stairway to the Mind*, New York: Springer-Verlag.

Searle, J. R. 1983 *Intentionality*, Cambridge: Cambridge University Press.

Searle, J. R. 1992 *The Rediscovery of the Mind*, Cambridge, Mass.: The MIT Press.

Searle, J. R. 2001 *Rationality in Action*, Cambridge, Mass.: The MIT Press.

Varela, F. J., Thompson, E., and Rosch, E. 1991 *The Embodied Mind: Cognitive Science and Human Experience*, Cambridge, Mass: The MIT Press.

Reasons, Truthmakers and Truth Grounds

HERBERT HOCHBERG

Commenting on Plato's remark, "When we assert *not-being* it should seem, what we assert is not the *contrary* of *being*, but only something *other*," Taylor wrote: "All that is necessarily implied by the proposition 'x is not y' is that x is something *different* from anything which is y." (Taylor, 257b, 164–165). This suggests reading Plato as either saying:

(i) x is not-F =df For every f, if x is f then f is different from F,

or

(ii) x is not-F =df For every y, if y is F then y is different from x.

Some Plato scholars consider such readings as problematic, and, in his introduction, Taylor states that *confined* to statements about *forms*, like 'Flying is not Sitting', one can take such negative statements as simple statements of difference. Applied to statements made about a concrete individual, he takes Plato to construe x being not-F in terms of x being something other than what *we call* 'F', and he takes that to involve F belonging to a family of properties that are incompatible. Somewhat later Owen also suggested reading Plato along the above lines:

> 'Theaetetus flies' says what-is-not about Theaetetus because what it says of him, viz. ... flies, is different from all the predicates he does have – or, in the locution that 263b echoes from 256e, different from 'the / many things that are with respect to him'. (Owen, 114–115)

Owen continues, in a footnote:

> This remains the simplest interpretation, requiring no shift in the sense of 'different' such as is sometimes found in 257b. To be sure, if it were taken as a rule for verifying or falsifying statements it would make falsification an interminable business, but this is not its function. If X is not beautiful, all X's predicates fall into the class different-from-beautiful introduced at 257d4–e11. (Owen, 115).

Owen thus takes not-F to be understood in terms of difference (≠) and sameness (=) in the sense of (i) rather than (ii), Here we can take = and ≠ as relations, ignoring questions about "relational" forms as well as obvious questions and problems raised by taking both = and ≠ as basic, instead of taking either negation (¬) and one of = or ≠ as basic. In one

B. Brogaard, B. Smith (Hg.), *Rationality and Irrationality / Rationalität und Irrationalität*, S. 152–171.

sense, given ¬, ≠, and =, it does not matter much which two are used to dispense with the third, given that two are required. (That is, forgetting questions about how negation (or difference) occurs in other contexts, there is the simple point that given that one requires at least two basic notions for handling simple negations of "atomic" statements, it matters not which two one prefers). But even allowing that that makes no difference, one cannot, from = and ≠, arrive at the appropriate equivalences '¬(f=g) iff f≠g', '¬(f≠g) iff f=g', and '¬¬Fx iff Fx' required of any attempt to dispense with one of the three in terms of the other two – whereas one easily does so with ¬ and either = or ≠. (This assumes the use of standard rules and not the addition of special axioms governing = and ≠.) One simply gets involved in unending sequences of iterations of negation and more and more complex contexts. Thus one has to adopt such patterns (or others they can be derived from) as basic "rules" or "axiom schema" governing negation. As these matters are not my main concern here, I just leave it at that and simply note that it is not merely a question of making falsification "interminable," as Owen puts it.

Bradley, Bosanquet and R. Demos attempted something along the lines of treating the negated property as belonging to a family of properties which are incompatible. What they did is reminiscent of some of Plato's remarks about "the large," "the small" and "the equal" – that the "not-large" can be the small *or* the equal. But Owen takes 'not-F' to be understood in terms of all the attributes that are "different from" F, rather than in terms of some (or one) that are (is) incompatible with F.

> If flying 'is not with respect to Theaetetus', the non-identity holds now between flying and any and all of the attributes which do belong to Theaetetus, which 'are' for him ... And thus he prepared his ground for contradicting the assumption used ... to extract contradictions from falsehood. The man who speaks or thinks falsely does after all and without paradox ascribe being to what is not, or not-being to what is: he counts among X's attributes one which 'is not with respect to X', i.e. which differs from any of X's attributes; or he counts among attributes of the second class one which 'is with respect to X'. (Owen, 131).

While Bradley and Bosanquet construed '¬Fx' in terms of '(∃Ø)(Øx & Ø is incompatible with F)', Demos took '¬p' in terms of '(∃q)(q is true & q is incompatible with (in opposition to) p)', ignoring meta-linguistic worries about 'is true'. In the *Logical Atomism* essays Russell argued that this simply gives us another form of negative fact – incompatibility facts – and took incompatibility in terms of the Scheffer stroke function. A year later, in "On Propositions: What They Are and How They Mean," he took Demos' view to generate an unending series, much like that involved with negation above, by considering how to state 'p and q are not both true'. (Russell, 1956a, 288–289.) Russell's point was that Demos was holding that propositions in "opposition" could not both be true, but yet he could not take that as an explication of 'opposition', since it involved using 'not'. Hence, Demos must apply his analysis, in terms of 'opposition', to that use of 'not'.

Russell argued for negative facts in both the logical atomism lectures and the 1919 essay. But in 1925 in the second edition of *Principia* another pattern emerged.

Given all true atomic propositions, together with the fact that they are all, every other

> true proposition can theoretically be deduced by logical methods. That is to say, the apparatus of crude fact required in proofs can all be condensed into true propositions together with the fact that every true atomic proposition is one of the following: (here the list should follow). If used this method would presumably involve an infinite enumeration, since it seems natural to suppose that the number of true atomic propositions is infinite, though this should not be regarded as certain. (Whitehead and Russell, 1950, v. 1, xv.)

Taking this in a context where atomic facts are the truth grounds for atomic propositions, and Russell's proceeding to construe all molecular compounds, including negations, as Scheffer stroke functions of atomic propositions, one can take him to have revived the idea that is implicit in the way Plato was read above. Only Russell does not attempt to "define" negation or replace negations by universal generalizations, since a basic (Scheffer) stroke function is employed in the second edition. Rather, what he says can be understood as an attempt is to avoid molecular facts, including negative facts, by appeal to atomic facts and a general fact that the atomic facts are all the atomic facts. That can be taken as the import of what he says even though he speaks of a general fact about a set (list) of true atomic *propositions* comprising all such truths, and not of a general fact about a set of atomic *facts*. I make that inference since Russell took negative facts, when he advocated such things, in 1918 and 1919, to ground the falsity of false atomic propositions and the truth of true negations of atomic propositions. In the 1925 passage true negations of atomic propositions are taken to be true since they are logical consequences of the statement expressing the general fact and the appropriate list of true atomic propositions. This would furnish the same sort of basis for his rejection of conjunctive facts in 1918, since the facts that grounded the truth of the conjuncts sufficed for that of the conjunctive statement without an additional conjunctive fact, as the truth of the conjunction was a logical consequence of the truth of the conjuncts.

Russell's idea has been revived in recent years by D. M. Armstrong and others. (Armstrong, 1997, 196–201; Simons, 1992; Hochberg, 1992). Armstrong appeals to a totality of facts and a special primitive relation of totality between such a totality of facts and a property, in this case the property of *being a* (first order) *fact*. He recognizes conjunctive facts, as mereological sums, in addition to atomic facts, in such a totality. Besides such a totality or "sum" of first order facts, there is a "meta-fact" that all the elements of the totality are all the first order facts. Simons takes there to be what amounts to a maximal class of atomic facts as truth makers, while I took there to be a domain of (class of all) atomic facts, with such a class providing the ontological ground for true negations of atomic sentences. A negated atomic sentence is true because the purported atomic fact is not in the class. The rationale for such a move is that it is not a fact that such an excluded atomic fact is not in the class, since classes alone provide the truth grounds for statements of class membership, like 'a $\in$ {a, b}'. (This will be elaborated on below.) Such attempts to use classes to avoid negative facts fail, but there is a sense in which Plato was right.

We can avoid conjunctive facts of atomic components by taking conjunctions of atomic sentences to be true in virtue of the atomic facts that ground the component conjuncts. Negation is another matter. First, there is no standard logical rule for negation corresponding to 'p, q $\vdash$ p & q'. Second, an evaluation assigns T or F, but not both, to the

atomic sentences, and thus the very mechanisms of the truth table assignments embody the familiar "laws" of non-contradiction and excluded middle and, thereby, *negation*. Thus the truth table for '¬' does not explicate or "define" negation. It simply interprets the sign '¬' as the sign for negation. Negation is already used in the very formation and use of the truth tables. One issue that then arises is whether atomic statements like '¬Fa' can be taken to be true in virtue of the atomic facts together with the general claim that they are all the atomic facts – if the negative statement can be derived from statements not involving a negation. A second issue that arises concerns what is to count as a negation. And, a third and crucial issue is whether true negations of atomic statements require an ontological ground of truth, or, as some would put it, a truthmaker at all – perhaps something that would be arrived at by "default," as Simons puts it – or whether one can viably speak, without ontological import, in terms of such statements being true in virtue of the "absence" or "lack" or non-existence of certain facts. My focusing on the issue of negative facts is not to presuppose that all ontological grounds or truthmakers are facts.

1. Deriving 'is-not' from 'is'

If one takes '¬Fa' to be "made true," to use Russell's phrase, by the general fact: (p)(p≠a's being F) – with 'p' as a variable over (atomic) facts and the formula is understood to state that no fact is a's being F – an obvious problem arises that is easily seen by limiting discussion to a small world (model), say with two atomic facts Ga and Fb. It is easily seen if one then uses '(p)(p=Ga ∨ p=Fb)' in place of '(p)(p≠a's being F)' to express the general fact that supposedly grounds the truth of '¬Fa'. Just as 'There are only two particulars' can be construed in terms of '(∃x)(∃y)(x≠y & (z)(z=x ∨ z=y))', 'a and b are the only particulars' can be construed in terms of '(x)(x=a ∨ x=b)'. That entails, in turn, with an additional name 'c', that '(∃x)(x=c)' entails '(c=a ∨ c=b)'. But as '(x)(x=a ∨ x=b)' does not entail '¬(∃x)(x=c)', neither does '(p)(p=Ga ∨ p=Fb)' entail either '¬(∃p)(p=Fa)' or ''¬Fa' is true'. All, of relevance, that follows is that '(∃p)(p=Fa)' entails '(Fa=Ga ∨ Fa=Fb)' – i. e. that '(Fa≠Ga & Fa≠Fb)' entails '¬(∃p)(p=Fa)'. Moreover, even getting so far *assumes* that we can instantiate to 'Fa', as we assumed about 'c' above. This, by itself, poses an insurmountable problem for the attempt to avoid negative facts if one takes 'Fa' to represent a state of affairs. For clearly it is a non-existent state of affairs, and it will hardly do to recognize such states of affairs in avoiding negative facts. That aside for the moment, stating that the truthmaker for '¬ Fa' is the non-existence of Fa, on the above pattern, involves 'Fa≠Ga' and 'Fa≠Fb'. This raises the question of what grounds the truth of the diversities, 'Fa≠Ga' and 'Fa≠Fb'. One cannot appeal here, as we will see some do in other cases, to the existence of diverse entities to ground a truth of diversity – that 'a≠b' is true in virtue of a and b, or the mereological sum a + b, and not in virtue of a fact of diversity, that a≠b. But, arguments about that aside, that cannot work here, as Fa does not exist. We will return to the issue about the truth grounds of diversities in a more general consideration of truthmakers below. Here we need only note that one cannot appeal to any view that, explicitly or implicitly, makes use of statements like 'Fa≠Ga' in grounding the truth of '¬Fa'.

The assumed instantiation to 'Fa', as we noted, raises the problem posed by non-exis-

tent states of affairs or Meinong's nonsubsistent objectives, since Fa does not exist and is not in the domain of the quantifier. Thus a question arises as to how one can even properly state that 'Fa≠Ga'. To consider it we must take up the question of the proper way of representing facts. Moore and Russell took facts as what "makes" or "renders" or "directly proves" a proposition true – a "truthmaker" or correspondence theory of truth – shortly after the turn of the century. In his version of the theory, set forth in lectures of 1910–11, Moore had faced a problem that was later put cryptically in *Tractatus* 4. 022 In Wittgenstein's terms, an atomic statement, or a "thought" that a is F, represents a situation – *shows* its sense – whether or not it is true, and *states* that it obtains. Showing or *representing* is a relation between a statement or thought, on the one hand, and a *possible* state of affairs on the other. This relation obtains whether or not the represented situation does, since the thought must have the same "sense" whether it is true or not. A statement cannot simply be taken to correspond to an existent state of affairs, as Plato noted in the *Theaetetus* and *Sophist*. Thus a problem arises for a correspondence theory of truth. (Russell, 1956, 182–183.) Armstrong appears oblivious to the issue and dodges it by speaking of "making true" and of the relation as "internal." Moore had dodged the issue by speaking of "directly proves" and of obtaining the name of the fact, whose "being" directly proved the truth of "the belief that-p," by subsituting the term 'fact' for the term 'belief' in the *name* of the belief. Both Armstrong and Moore take 'Fa' to be true if and only if a fact makes it true, but both note that it is not the *existence* of any fact that makes 'Fa' true. As Armstrong thinks of it, for example, the existence of a certain fact "necessitates" the truth of 'Fa' – the fact that a is F. Putting it this way, using the clause 'that a is F', points to the obvious problem – the implicit recognition of states of affairs that are mere possibilities, or have "no being" in Moore's terms, since what makes 'Fa' true is the existence or the *obtaining* of the state of affairs that 'Fa' represents. Thus "correspondence" is ambiguous. In one sense 'Fa' corresponds to a state of affairs whether it is true or not; in another sense, if true, it corresponds to an existent fact – its truth maker.

As Armstrong speaks of the relation between a statement and its truth maker being "internal," Searle takes the same relation between a proposition and its "conditions of satisfaction" to be "intrinsic." Both ways of putting matters simply amount to holding that there is no relation and no problem as it is the existence of the truth bearer, 'Fa', and its truthmaker that makes "'Fa' is made true by the fact that a is F" or "'Fa' has as its condition of satisfaction the fact that a is F" true. (Armstrong, 1997, 129; Searle, 1983, 4–17) Such claims pose an obvious and familiar problem. Armstrong and Searle both assume that as atomic sentences have a meaning, without reifying meanings, Meinongian non-subsistent objectives or possible, but not actual, facts are avoided. But, like Wittgenstein, they simply pack the represented possible facts, as essences (internal properties), into the terms, via claims about the connection between 'Fa' (or the belief or proposition that a is F) and the fact that a is F being an "intrinsic" or "internal" connection. Searle thinks that he resolves the problem by distinguishing between "intentional-with-a-t" and "intensional-with-an-s." Thus "the belief that a is F" is intrinsically related to a condition of satisfaction, even though the latter does not exist, because that simply means it does not have an "extension" but retains its intentional feature – its being intrinsically related (to what doesn't exist). Searle's purported solution is empty as it simply restates the problem. (Hochberg, 1999, 120–127)

The problem raised concerns the truth grounds for statements of modality as an atomic sentence is implicitly taken to represent a possible state of affairs. Armstrong has sought to develop a combinatorial account of possibility along what he takes to be Wittgenstein's *Tractarian* lines to avoid possible facts but account for the truth that it is possible that a is F. What he did was claim to dispense with possibilities in terms of actual facts. But like Moore, Wittgenstein and Searle, he really recognized non-actual states of affairs represented by atomic sentences. For he thought he disposed of mere possibilities by taking the mereological sum of the constituents, which is actual, as the purported truth maker for the claim that a state of affairs is possible. But it is easy to see that by asserting that the constituents determine the possible situations and declaring that doing that only recognizes the constituents, he merely takes the natures of the terms to determine a possibility of combination. And that simply takes the sentential juxtaposition of 'F' and 'a' in 'Fa' to represent a possible situation. He thinks that he takes the truthmakers for modal statements to be actual entities. The mereological sum F+a, as a truthmaker, supposedly grounds the truth of '◇Fa' (or "'Fa' expresses a possibility"), a truth bearer, and '◇aRb' and '◇bRa', as truth bearers, are made true by the mereological sum a+R+b, their truthmaker. But what he does is disguise possible facts as the *quiddities* or *natures* of the objects, properties, and relations. For he must hold that the modal truth bearers are true since the elements of the mereological sums in their truth makers *can* combine in certain ways. Thus, like Wittgenstein, he let the natures of the elements carry the possibilities of combination and determine the possible facts. Thus possible facts are packed into the essential properties of universals and particulars, since such essential properties determine the possibilities of combination. A particular is what *can* combine with a monadic first order univeral to form an atomic fact and with other particulars and appropriate relational universals to form n-adic facts. Likewise, what is involved in the nature of monadic universality, dyadic universality, *etc.* are the possibilities of combination. It is not simply the mereological sum a+F that suffices to ground 'Fa' being well formed and expressing a possibility. He noted a problem with his combinatorial acutalism when he held that an assumed totality of particulars, a+b, was the truth maker for '◇(∃x)(x≠a & x≠b)', where talk of "deflation" is transparently inadequate. This led him to speak of a possible "alien" being relegated to an "outer realm of possibility" which was justified by *haecceitism* not being involved, as we do not talk about any specific particular. This obviously will not do.

Recently, to meet objections to his views, Armstrong has proposed a new solution to the ontological problem posed by modalities. He suggests that Fb's being *contingent* suffices to ground the truth of '◇¬ Fb'. (Armstrong, 2000, 155–159) It is the contingency of b's being F that is the truthmaker for the modal sentence. One thing that is obviously wrong with his new approach is that it faces the same problem as his earlier view. We see that immediately if, instead of simply dealing with the modal sentence, we ask what grounds 'Fb' being well formed – i. e. expressing a possibility, in an obvious sense of that term. He does not face the question because he focuses on the sentence '◇¬ Fb', which already makes use of the sentence 'Fb', and, hence, assumes it represents a possible state of affairs. Another problem is that his use of 'contingent' remains vague. If the modal sentence is to be true because the sentence 'Fb' is contingently true, then he returns to a simplified version of Carnap's notions of L-true and L-false, with a standard use of 'logical truth' whereby neither 'Fb' nor '¬Fb' are logical truths or falsehoods. If he is taking the

fact that b is F to embody a modality of contingency, then he has merely compounded the problem. For not only has he introduced a basic modality, as a mode of facts, but he must now ground the connections between facts being contingent and truths about possibilities, since one would think that 'p is contingent' is understood in terms of necessity and/or possibility, as Carnap once noted: "... 'p' is contingent means 'p' is neither necessary nor impossible ...".

Let 'C' express 'being contingent' and be a sign that combines with sentences, so that we have 'C(Fb)' as a well-formed pattern. What Armstrong must do is either introduce a host of axioms, or Carnapian style semantical rules, so that we have 'C(Fb) ⊃ ◇¬ Fb', or problematically seek to define the other modal concepts in terms of 'C' and '¬'. He probably does not note that he must do such things as he takes all facts to be contingent, and, as he denies that there are any necessary facts, he speaks of Fb being a fact as amounting to characterizing it as contingent. But if he introduces a new basic modality of contingency, he must then either make the modality of contingency an internal or necessary property of facts or take it as a basic mode that does not form a further contingent fact consisting of the fact Fb *exemplifying* the mode of contingency. Yet, if C is an internal property, then, by Armstrong's notion of an internal property, it supervenes, and hence the truthmaker for 'C(Fb)' is simply the fact Fb, as what is supervenient is no "addition to being." Thus he would end up simply declaring that since the fact Fb is the truthmaker for 'C(Fb)', and there is no further state of affairs involved, Fb is the truthmaker for '◇¬ Fb', as the latter is true in virtue of the contingency of Fb. So he must take C(Fb) to be a further fact, one that is necessary, as he leans towards accepting the modality C as basic and holding that 'C(Fb)' is a necessary truth. He does not avoid such a necessity by holding that necessary truths have as truthmakers "the possibility of the existence of their terms." (Armstrong, 2000, 158.) Hence, the (possibility of the) existence of C and Fb ground the truth of 'C(Fb)'. Such a use of 'necessary truth' is puzzling, while his use of 'possibility' is both unexplicated and circular, as he not only faces the problems about 'possibility' we noted earlier, but he now uses the notion of *possibility* to explicate 'C(Fb)' being a *necessary* truth. Yet the latter is used to explicate '◇' in '◇¬ Fb'.

Moore's 1910-11 view and Armstrong's and Searle's later versions rely on an implicit referential connection between sentences, or beliefs they express, and a possible fact. Russell had indicated a way of avoiding doing that in 1905 that he tried to develop in 1913 but abandoned under Wittgenstein's influence. (Russell, 1984, 144–145) In 1905 he suggested, and in 1913 explicitly held, that "'Fa' denotes the fact that-Fa" be replaced by:

(R) 'Fa' is true if and only if the fact *consisting of* a and the property F exists,

where the implicit existential quantifier ranges over *existent* facts, not possibilities, as Russell's existential quantifier for particulars ranges over existent individuals. (R) makes no use of a representational relation – *designates* or *refers to* – connecting a sentence to a complex, a purported state of affairs. Instead, it employs a Russellian description that purports to denote an existent complex and is more explicitly rendered by:

(R*) 'Fa' is true iff Fa iff E! (ι p)((a is a term in p) & (F is the attribute in p) & (Øx is the form of p)),

where 'p' is a variable ranging over (existent) atomic facts.

(R*) is a tripartite biconditional since it is an interpretation, but *not* designation, rule that also specifies a truth ground – and hence plays the role of a Carnapian rule of truth as well as providing an interpretation for the sentential juxtaposition of 'F' and 'a' in 'Fa'. Russell's theory of definite descriptions allows us to avoid mere possibilities or Meinongian objectives, which Armstrong and Searle do not manage to do. Moreover, it is easy to see that such *formal* or *internal* relations (*is a term in*) are immune to regresses, which is why they can be taken to be formal or logical or internal. With 'T', 'A' and 'IN' for such formal relations (is a term in, is the attribute in, and is the form of, respectively), to specify the fact that would be a term of a further relational fact requires describing it as (ι p)(T(a, p) & A(F, p) & IN(Øx, p)). That a is a term in such a fact is then expressed by: a is a term in (ι p)(T(a, p) & A(F, p) & IN(Øx, p)). As this simply says that the fact exists, on Russell's analysis of definite descriptions, it trivially follows that the fact exists if and only if a is a term of it. This *shows* that such formal or *internal* relations do not give rise to further facts – which is why they are formal (internal). It also shows that facts are basic, but neither simple nor perspicuously labeled by names. The point applies to the attempt to label any complex and lies behind Russell's view that only simples are perspicuously referred to directly ("tagged" as some now say). But such reference involves a basic intentional or designation relation. In line with (R*), a negated existential quantification, '¬E! (ι p)(T(a, p) & A(F, p) & IN(Øx, p))', can express the non-existence of Fa. The issue is then whether such a pattern avoids negative facts.

The claim would be that by recognizing the class of atomic facts, we can specify a truth ground in terms of the assertion, or denial, that there is a fact with certain constituents and of a certain form. One might then argue that such a view enables the correspondence theorist to accommodate true negations wihout recognizing negative facts, for one is recognizing an ontological basis for taking such statements to be true: the set of atomic facts, given that we have recognized such a set as the correlate of the sentential variables – as the set of atomic facts that constitutes the domain. Thus it is tempting to argue that it is no more a further fact that no such member of the set exists than it is a further fact that such a fact does exist, if the sentence 'Fa' is true. That is, where the sentence 'Fa' is true, a certain fact is taken to exist. There is no ground for holding that there is then an additional fact, the fact that the atomic fact exists. And there is no more need to recognize a fact that an atomic fact does not exist when it does not than there is to recognize a fact that an atomic fact does exist when it does. This is reinforced by our conception of what makes a statement of class membership true. Recall that it is not a relational fact involving class membership, but the class itself.

If the appeal to a set of facts taken as the domain of atomic facts or to a totality of facts is viable we can avoid negative facts. But there is a simple argument against such a view, which is obscured by Armstrong's presentation, as well as by my earlier argument appealing to classes, by Simon's version of the view, and by Russell's original variant of it. We cannot say, where '¬ Fa' is true, that *the fact* Fa does not belong to the totality or *is not;* or, as Russell would express it – that the fact Fa is not identical with any of the existent atomic facts by implicitly using the the expression '*the fact* Fa' to denote the fact *that* a is F. Rather, we must describe that fact. This means that we must claim that the described fact is not one of the existent facts, which is to say that the described fact does not

exist. We are thus back to the question of what is the truthmaker for such a claim: '¬E! (ι p)(T(a, p) & A(F, p) & IN(Øx, p))'. An obvious point then emerges that Russell's theory of descriptions makes transparent. To claim that '¬Fa' is true since the described fact does not belong to the totality or class of atomic facts is simply to repeat '¬E! (ι p)(T(a, p) & A(F, p) & IN(Øx, p))'! Thus we return to the issue of whether such a negative truth requires a truthmaker.

Given what we may call "the class form," {...}, we can construe classes, ontologically, as composed of such a form and the members of the class. Taking classes in this way, we can easily see why a truth like 'a ∈ {a, b}' is trivial, while one like 'a ∈ {x| Fx}' need not be. The first sentence says that a is an element in a class, but that class is implicitly described as resulting from the class form combining with a and b. The second sentence states that a is an element of a class that is described, without reference to any specific elements, but as the complex that results from the class form combining with the objects that exemplify F. Thus the first statement is true simply on the basis of our assuming that classes exist, given elements and the class form. This is another way of putting the point that the ground of truth for 'a ∈ {a, b}' is simply the class itself, not a fact involving a purported relation, ∈, holding between the class and an element of it. 'a ∈ {x| Fx}', by contrast, is a disguised conjunction. For what it states is that a is an element of a class, which exists by our assuming classes exist, and which exemplifies the property F. It is assumed that '{x| Fx}' is a non-trivial definite description of a class, where descriptions like '(ι α) (x)(x ∈ α ≡ (x=a ∨ x=b))' are trivial, and the elements are taken to be indicated by simple indicatives and not by non-trivial definite descriptions, such as '(ι x) (Gx)'. These brief comments about classes apply directly to our problem concerning negative facts. For the claim that (ι p)(T(a, p) & A(F, p) & IN(Øx, p)) does not belong to the class of atomic facts involves two descriptions; one of the particular fact and one of the class, taken as the domain of atomic facts. Thus we deal with a statement on the order of '¬(ι x) Gx ∈ {x| Fx}' and not one like '¬a ∈ {a, b}'. This means that we cannot simply appeal to a class or domain as the truth ground for '¬E! (ι p)(T(a, p) & A(F, p) & IN(Øx, p))' or '¬Fa'. Notice, too, that '¬c ∈ {a, b}' poses a problem, unless it is understood that 'a≠c & b≠c' – or that different names (as types, not tokens) represent diverse things. Hence, in the case of statements like '¬c ∈ {a, b}', the class *and* the facts of diversity, a≠c and b≠c, suffice as truth grounds. There are no facts of membership, positive or negative, as truth grounds for statements of class membership (positive or negative) – the classes, if taken as entities, suffice, along with diversities in the case of negations. The basic problem faced by the various attempts to use a generality to dispense with negative facts is seen if we consider:

(N) (q)(q ≠ (ι p)(T(a, p) & A(F, p) & IN(Øx, p)).

Where the variables 'q' and 'p' range over existent atomic facts, this serves the same purpose as a "list" or a corresponding disjunction, and avoids questions about "infinite" lists or disjunctions. If '≠' is a primitive sign, as *diversity* (as opposed to *identity*) is taken as phenomenologically basic, then (N) becomes:

(N′) ($\exists_u$p)((T(a, p) & A(F, p) & IN(Øx, p)) & (q)(q≠p))

using the subscript 'u' to indicate "uniqueness" of *the* p in place of a uniqueness clause in the expansion of the description. This will obviously not do as an expression of the truth ground for '¬Fa', since there must then be the fact of a's being F and yet it must be diverse from every (existent) fact, which is absurd. This would force one into recognizing existent and non-existent (merely possible) states of affairs with corresponding variables ranging over such respective domains, as in some so-called "free" intensional logics.

If we reject diversity as basic and treat '≠' in terms of '¬' and '=' and take (N) as

(N″) (q)¬($\exists_u$p)((T(a, p) & A(F, p) & IN(Øx, p)) & (q=p)),

we are back to the problematic use of a negated existential claim: for any fact it is not the case that there is a fact consisting of a and F and of the form Øx identical with it. The same kind of point involving the need for appealing to a negated existential claim or the use of '≠' arises in the case of an argument Smith uses for illustrating the existence of truths without truth makers (Smith, 1999). Take your favorite non-existent object, his is apparently Ba'al. Then nothing in the world makes it true that such a "thing" does not exist. That claim resolves nothing. Treat the apparent "name" as a Russellian definite description. We are then back to a negated existential claim or a universal claim holding that every thing (particular, in this case) is such that it is diverse from (or not identical to) the described (but non-existent) particular. That simply takes us back to statements that raise, not resolve, the problem.

All one can derive from the generalization '(p)(p=Ga ∨ p=Fb)', allowing for the problematic instantiation (or, what amounts to the same thing, assuming in addition '(∃p) (p=Fa)'), is '(Fa=Ga ∨ Fa=Fb)'. This does not get us to '¬Fa' or "'¬Fa' is true." Thus, Russell, strictly speaking, was wrong in his claim, unless he understood it, as he may well have done, to involve taking all atomic sentences *not* on the list to be false or their negations to be true, which can be understood as recognizing negative facts.

But there is an alternative way of attempting to provide a ground for the truth of '¬Fa'. Consider the following derivation, with 'p' and 'q' ranging over atomic facts:

(D) 1. a≠b
2. F≠G
3. (p)(p=Fb ∨ p=Ga)
4. (p)(q)[(IN(Øx, p) & IN(Øx, q)) ⊃ {(x)(y)(f)(g)((x=y) & (f=g) & T(x, p) & A(f, p) &T(y, q) & A(g, q)) ≡ p=q}] – Monadic (first order) atomic facts are the same iff their constituents are the same.
5. (∃x, y, f, g)(x=a & y=b & f=G & g=F) ∴ ¬Fa.

With 'Fb', 'Fa' and 'Ga' understood as abbreviating the appropriate descriptions of facts indicated earlier ('¬Fa' is then a negated existential claim), it is a valid argument. For, assuming an existential statement, using the description that 'Fa' would abbreviate, we can instantiate (3) to (6) 'Fa=Fb ∨ Fa =Ga'. But, by (4), (6) is false, so we arrive at '¬Fa', i. e. '¬E! (ι p)(T(a, p) & A(F, p) & IN(Øx, p))'. We have then "grounded" the truth of '¬Fa' without appealing to a negative fact – a's not being F – by either taking '¬E!(ι p)(T(a, p) & A(F, p) & IN(Øx, p))' to express there being such a fact or introducing '(ι p)(T(a, p) &

A(F, p) & IN(¬Øx, p))' to describe a fact. This points to (1) and (2) – the facts of "difference." Are they "negative" facts? Are they really entities, i. e. are there such facts, or do the diverse entities suffice as truth grounds? This raises the issue of whether a≠b, or ¬(a=b), taken as a fact, is a negative fact, as well as questions about whether ≠ (difference) is a basic relation, and not to be understood simply in terms of '¬(...=---)', and about whether the truth grounds for true statements of difference are simply the diverse entities or if the fact that they are diverse is required.

Armstrong, Simons, Mulligan and Smith claim that a *and* b suffice as "truthmakers" for 'a≠b' and that such a true statement of diversity holds without the need to recognize the relation ≠ as a constituent of a fact of diversity. The claim is sometimes based on diversity being "internal" and, hence, as Armstrong puts it, merely supervenient. Thus it is not the terms standing in such a relation that provides the truth ground for the corresponding relational statement, but the *terms* alone (sometimes identified with their mereological sum). But that cannot be, as Plato, in his way, recognized long ago when he took difference as a basic form. For it is the existence of a *and of* b, where the 'and' carries the sense that a is diverse from b. Simply put, (1), in (D), does not follow from (5). It is not just that we have the list of existents 'a, b', but that the list is a list of diverse entities. That a is diverse from b is a basic matter of fact. Armstrong implicitly acknowledges the failure of his own view, when he accepts a suggestion D. Lewis made and adopts a special notation for (necessarily) symmetrical relations. Thus he replaces 'a≠b' by 'they (a and b) are different'. The use of the clause 'they (a and b)' points to the problem, for Armstrong can be taken to be using it to do one of two things. First, he can be seen as implicitly introducing a new entity, a complex composed of the two objects, and as using diversity as a monadic property of the complex *they* form. This is why Gustav Bergmann's explicit recognition of *diversities* as entities does not suffice (Bergmann, 1992, 101 ff., Hochberg, 1999, 96–99). But even with such a new complex entity, Armstrong, like Bergmann, would clearly beg the question, for to form such a complex there must be *diverse* elements. Second, and what he explicitly does say, Armstrong can hold, as Mulligan did during the discussion, that we simply have a mereological sum, a + b, as "truth maker." But that also begs the question, for one takes the "sum" to have proper parts, and there can only be proper parts of a mereological sum if there are diverse elements. We then still take *having proper parts* as a monadic property of the complex, the sum in this case, as is clear from Armstrong's formulation 'they (a and b) are different' where "being different" is attributed to "they" – (a and b). This emphasizes that one has not avoided the "form" or relation of diversity. All one does is turn it into an odd monadic property of a complex composed of diverse elements.

Second, if ≠ is construed in terms of ¬ and = then it is also clear that facts of diversity are negative facts. But one need not argue about that. Such facts are not facts like ¬Fa would seem to be. Thus the proponents and opponents of negative facts each have a point. Third, and crucially, if ≠ is basic there is a clear sense in which we avoid negative facts altogether and, hence, a sense in which Plato was right after all. I think, phenomenologically speaking, it is clear that diversity is basic. One is presented with or directly aware of such facts of diversity, but never with the fact that something is self-identical. (Segelberg, 1999, 38–41; Hochberg, 1999, 50–54). Yet there is an *apparent* problem. If diversity is taken as basic then '≠' should be a primitive sign used to define '='. But this appears to create problems for the use of definite descriptions – problems that I can only indicate and not probe further here. Consider '(q)(q ≠ (ı p)(T(a, p) & A(F, p) & IN(Øx, p))'. One can, with '≠' as primitive, only expand the description so that we have an existential claim – either explicitly or embedded, depending on whether we take the scope to include the universal quantifier or not – which is hardly what is intended. Just how one can satisfactorily resolve this matter I cannot take up here.

There is a related, simple, even trivial point to note in connection with the rejection of negative facts. If the totality of positive atomic facts will suffice, along with a general "meta-fact," so will a totality of negative facts and an appropriate generalization. One can also take '≠' as basic, rather than '=', and express the appropriate universal generalization as a universally quantified conjunction employing '≠', for diversity, in its atoms, rather than '='. Taking the existential quantifier as basic, one can then construe the appropriate generalization as a negated existential statement. Thus there is no reason to prefer a generalized meta-fact that is a positive fact and hence claim that one only recognizes positive facts. We can just as easily have all facts as negative, and Plato could have a new trio of basic forms: negation, difference and existence. Moreover, as noted, there is a phenomenological argument for diversity, not sameness, being basic, as one clearly forces matters to claim that one apprehends that something is one and the same with itself. It is rather a matter of a purported logical principle that every entity is the same as itself. This is obviously why philosophers have sometimes sought to construe 'exists' or 'self-identity' in terms of the other. What I just called a trivial point has an additional aspect, as it concerns a philosopher like Armstrong who accepts conjunctive but not disjunctive facts. For the universal "meta-fact" that is needed is a universal disjunction, if one is to avoid using negations in any way. He needs a fact corresponding to a statement like '(p)(p=Ga ∨ p=Fb)'. By contrast '¬(∃p)(p≠Ga & p≠Fb)' allows him the appearance of using a conjunctive form. Perhaps the basic point to note is that, in connection with (D), one cannot take any one fact or entity as a so-called 'truthmaker' of '¬Fa'. There are facts of diversity, a generality, and the logical patterns and forms involved. These, together, provide *the reason* '¬Fa' is true and indicate what is involved, ontologically speaking.

We are now in a position to provide reasons and, if needed, grounds for '◇¬ Fb' being true and 'Fa' being a well-formed formula, without non-existent possible facts. Some necessary truths will be so grounded in that they are based on the possession of a categorial property by an entity. Thus that a is a particular, given that a exists, is such a necessary truth. It is reasonable to take it to be a logical truth in the sense that it is reflected by the very categories and the formation rules of a schema. Thus in the context of a PM type schema one can say that 'b is a particular' is reflected by '(∃x)(x=b)' and 'Fb' being well-formed, while 'bF' is not. One might even hold that to say that 'b' can combine with 'F' to form an atomic sentence can be seen to be a more basic logical truth than the tautology 'Fb ∨ ¬Fb', as the former is embodied in the rules of the schema and not expressed by a sentence of the schema. To be a particular is to be something that *can* exemplify a property, and stand in relations, but that *cannot* be exemplified. Such basic modal notions (having nothing to do with modal logics) and the logical forms of exemplification (monadic, dyadic, etc.) are involved in categorizing particularity and universality. Nothing further is needed, and no further explication is to be had, but one must then recognize, as Armstrong does not, the non-supervenient categorial forms and an extended sense of

'logical necessity'. Other necessities and possibilities will be so in virtue of standard logic. Logic itself may be taken to be ontologically grounded by logical forms or logical "facts" reflected by logical truths, such as '(x)(∃f)fx', and logical relations among such forms that constitute more complex logical forms or facts, grounding more complex logical truths, such as '(x)(f)fx ⊃ (∃x)(f)fx'. The simple reason '◇(∃x)(x≠a & x≠b)' is true is that '(∃x)(∃y)(z)(x≠y & (z=x ∨ z=y))' is not a logical truth, and not that there is an existential fact exemplifying a primitive mode of contingency. If there is an ontological ground for such a possibility it will lie in the ground for logical truths being such. But does such a view fare better than Armstrong's and Smith's unexplicated use of modal notions (Smith's will be considered further below)? It clearly does by explicitly recognizing logical forms and the categorial properties based on them, distinguishing particulars from attributes (relations), that ground the modal concepts of logical possibility and/or necessity, which are represented by the structural categories of a PM type schema and have nothing to do with modal logics. This can be spelled out further. First, I have not implicitly taken 'Fa' to stand for a possible fact, as Armstrong and Searle do. We simply have a rule like (R*) and the "fact" that 'E!(ι p) (T(a, p) & A(F, p) & IN(Øx, p)) ∨ ¬E!(ι p)(T(a, p) & A(F, p) & IN(Øx, p))' is a logical truth. Second, we are all, knowingly or not, forced to recognize basic logical forms, like Øx, that ground predicative juxtaposition. There is no explicating *predication*, which is why so many try to replace it by merelogical relations, or some other pattern. Third, we must recognize the categorial distinctions that attempts to employ modal axioms about purported modal operators in functional modal logics already incorporate – the subject-predicate distinction. Such categorial distinctions are obviously not explicated by axioms about modal operators in such systems. Fourth, as argued above, sentence patterns like 'T(a, (ι p) (T(a, p) & A(F, p) & IN(Øx, p)))' are not relational statements of fact. No additional statements of fact are involved, and possible, but not actual states of affairs, are avoided by the use of definite descriptions of facts.

2. Truth Grounds

The unfortunate term 'truthmaker' has come into common usage, largely due to a now well known paper on the topic. (Mulligan, Simons, Smith, 1984.) I will ignore the term and instead start from the obvious truism that given a truth, there is a reason it is true. The question is what constitutes a reason. I am not speaking of reasons for holding something to be true – evidence as it were, and thus am not raising what some would take as an epistemological issue. Consider the case of negation that we just discussed. One furnishes a reason, in a clear sense, not by citing the existence of a negative fact, but by deriving the truth of '¬Fa' from a set of other truths (or purported truths or principles within a certain kind of logical framework) via the pattern in (D). If I was right to claim that facts of difference were appealed to then there is a sense in which the existence of such facts can be taken as among the reasons or grounds of the truth of '¬Fa'. So, in that sense, the issue of negative facts, regarding atomic facts, comes down to the issue about facts of difference. We will return to that shortly. For the moment contrast the appeal to (D) with the claim that no ground of truth is needed for '¬Fa' in that it is true by "default," as it were, since it

is the "absence" of a fact, that a is F, that is the "reason" the sentence is true. We thus have two fundamentally different kinds of "reasons" – the existence of facts and the derivation of sentences, on the one hand, and the "absence" or "lack" or "nonexistence" of facts on the other. One might even argue, as some do, that there are cases where there cannot be a ground for a truth being true. Consider the possibility that the universe is empty. (Both Armstrong and Smith have raised such a case in the context of the present issue.) The idea is that surely the existence of the empty universe would not be a ground of truth in case that possibility was realized. But this is misleading. First, it is one thing to ask about the truth ground of the possibility that the universe is empty. And even here that is quite vague, since one could be talking about the domain of objects being the empty domain or one could be talking about there not being a domain, i. e. not even the empty set. It would still not be clear what one then meant by 'the universe', with the implicit existential quantifier, in speaking of "*the* universe being empty" or, for that matter, what would be the bearer of truth. What would truth be ascribed to in such a case? It is all too nebulous to be a matter of serious concern. (see Hochberg, 1957)

Such extreme cases aside, talk of the "lack" or "nonexistence" of something being a ground of truth will not do. That becomes obvious, or should, if one considers what it amounts to by contrast with a pattern like (D). How do we express the "absence" or "lack" or "nonexistence" of the fact that a is F? It is obvious! By one of the formulations introduced earlier: '(q)(q ≠ (ι p)(T(a, p) & A(F, p) & IN(Øx, p))' or, simply, '¬E! (ι p)(T(a, p) & A(F, p) & IN(Øx, p))'. Then, with

T'¬Fa' ≡ ¬Fa ≡ ¬E! (ι p)(T(a, p) & A(F, p) & IN(Øx, p))

as the combined interpretation-truth rule for '¬Fa', we trivially derive '¬Fa' from either claim. But no truth ground or *reason* is given, in that the statement that the fact does not exist is, in a sense, derived from itself. The derivation in (D) is quite different, as no version of '¬Fa' occurs as a premise. To put it another way, to speak of a "lack" or "absence" or "nonexistence" of an atomic fact, in talking about the reason '¬Fa' is true, is to introduce a quite different kind of reason – the absence as well as the presence of facts. Thus one introduces a dichotomy that does exactly what the dichotomy of positive and negative facts does. All that differs is the way of expressing matters.

There is an odd point worth noting in connection with talk of being true in virtue of an absence, or lack or the non-existence of something. Consider a general truth, '(x)(Fx ⊃ Gx)' for example. Given that it is equivalent to the negation of an existential statement, one could take it to be true in virtue of the absence or lack of an appropriate fact or pair of facts, by the same sort of reasoning employed by those who appeal to such a notion of absence in the case of '¬Fa'. Thus, the absence of an object which was an F but *lacked* being a G (thus involving a double "lack") would suffice, on such a pattern. There would be no truth ground for such a general truth being true. This, in an obvious way, even out-does an extreme Humean approach to the truth of generalities and "laws." But it is an obvious consequence of the appeal to the absence of a truth ground for a statement being true as the reason its negation is true. ('⊃' could even be replaced by a *causal conditional* above.)

Where we have a truth, †, expressed by an atomic sentence, and a fact f, then f is a

ground of truth for † iff the statement that *f* exists entails the statement that † is true. I leave aside questions about 'b exists' or '(∃x)(x=b)' having b as a truth ground, as the context is a schema where all logically proper names and primitive predicates are interpreted as referring to particulars and properties. Likewise, I am not here concerned with statements other than atomic statements or negations of such and construe 'a≠b' as an atomic statement or the negation of one. Some set of statements S is the *reason* for a truthbearer †*, where †* is either an atomic statement or the negation of one, being true if S contains an existential statement that is a ground of †*'s truth or if (i) S contains statements that have truth grounds, (ii) all statements in S are true, and (iii) S entails †*. This avoids speaking of *necessitating*, in some ultimately unexplicated sense. Consider the idea that the existence of a *and* b – or (they, a *and* b) – necessitates the truth of 'a≠b'. Clearly, as noted, we do not have '(∃x)(x=a) & (∃x)(x=b)' entailing 'a≠b'. Thus we cannot appeal to ordinary logical entailment to dispense with facts of difference. (Though we can pack them into the use of 'and', which was the point of italicizing 'and' and using the parenthetical expression above.) Recognizing this, as we did in D earlier, we need only appeal to standard logical entailment. On such differences philosophical views stand and fall.

Regarding the notion of *necessitation* , which he will use, crucially, in a series of definitions to arrive at the definition of 'x makes p true', Smith suggests that:

> Necessitation is to be conceived as a real tie spanning the divide between ontology and logic. We define:
>
> DN xNp := E!x & (E!x ⊣ p),
>
> where p ⊣ q abbreviates ¬◇(p &¬q), and where p, q, ..., are schematic letters standing in for particular judgments ... and other candidate bearers of truth. (Smith, 1999.)

It is then easily seen where the problem lies. For to take the existence of a *and* b to necessitate 'a≠b' is to take 'a exists and b exists' to be true along with '((a exists & b exists) ⊣ a≠b)'. Clearly if '⊣' is read in terms of '(a exists & b exists) ⊃ a≠b' being a logical truth, i. e. in terms of '⊦', the claim is false, unless one presupposes 'a≠b', which is hardly enlightening. It is also problematic, since he speaks of a necessitation relation while 'necessitates' is a defined predicate. That aside, Smith does not think in terms of the above conditional being a standard logical truth, as he uses a primitive modal notion throughout – thus employing a notion of necessitation that is never explicated. (I do not consider taking a modified version of a modal calculus to govern the use of modal signs to provide either a philosophical explication of modal terms or an ontological analysis of modality.) As his use of 'entails' is not that of 'logically entails', Smith's opening abstract is quite misleading:

> On the one hand is the relation of necessitation, which holds between an object x and a judgment p when the existence of x entails the truth of p.

One expects standard logic. One gets "principles of necessitation" and the use of a primitive modal operator. Moreover, even treating ⊣ as a primitive necessary connection (ignoring his use of '◇' rather than '□'), he must still assume that a≠b. One gets absolutely

nowhere with the claim that it is *simply* a *and* b that furnishes the truth ground for 'a≠b'. Armstrong's claim that it is the mereological sum a+b that suffices, does not help. For, as noted earlier, one presupposes that a is a proper part of the sum or, to put it differently, that a+b≠a. There is no escaping the need to appeal to a diversity.

There is another problem. Taking '(E!x ⊣ p)' to be true in a relevant case, using some modal term (be it '◇' or '□') in an unexplicated sense, raises the issue of whether such truths require truthmakers. If not, what is the reason they are true? If so, we have a modal truth regarding x's existence (that x exists) and p, where p is a truth bearer of some kind – sentence, judgment, proposition, etc. Thus there is a primitive modality involved in connecting the the two terms. This raises a question as to whether we have modal facts grounding such true necessitations (perhaps necessary facts) involving a "real tie spanning the divide" between ontology and logic. If not, one can only invoke stipulative definitions and appeal to logical entailment, which does not work. Smith appears to be satisfied by a set of axioms governing his use of 'necessitates' that follow the lines of some modal calculus and axioms about '◇'. But this merely codifies, and neither explicates his use of 'necessitates' nor avoids a basic modal connection between truthmakers and truthbearers. Thus, no real explication of 'necessitates' or 'makes true' is offered, in spite of the long chain of defintions he introduces linking 'necessitates' with 'makes true'.

In some cases, for Smith as for Armstrong, the truthmaker that necessitates a judgment (proposition, statement) being true will be a fact. Thus the appropriate sign replacing the variable 'x' in 'E!x & E!x ⊣ p' will be a sign for a fact. Hence, Smith is forced to use a definite description of a fact or introduce purported names of facts and the possibilities that go along with them, as Carnap knowingly did and Armstrong and Searle unknowingly do. What sort of definite description is open to him? Consider the case of 'Fb'. However he takes the definite description, which, assume, is '∂', '∂' must be such that 'E!∂ ⊣ p' holds. Thus, even with an unexplicated modal notion, so that he does not have to show that 'E!∂' entails 'Fb', (which I believe is reason enough to reject his view), he must at least make it plausible that he can speak of a conditional being a necessary truth. Thus he will have to have introduce some sort of claim that plays the role of (R*), on the view presented earlier. But, then, for atomic sentences, and their negations, his discussion of 'makes true' and 'necessitates' is irrelevant. (R*) and the discussion of (D), along with a straight-forward concept of logical entailment, capture what is viably involved in talk of truthmakers and necessitation. Giving corresponding truth rules for complex propositions (statements, judgments) will do the same in those cases. Questions about whether objects, like b, are also truthmakers – for existential claims – are of no real import, as '(∃x)(x=b)' will either be used to define 'E!', as it is used with names, or the latter will be taken as a primitive predicate of existence. The problems involved in the latter I ignore, for, in either case, the statements will be true in a perspicuous schema where all primitive predicates and logically proper names (primitive zero level signs of the schema like 'a', 'b', *etc.*) will refer to properties (relations) and particulars, respectively. Whether one then says that it is b or b's existing that is the relevant ground of truth is of no import. In the case of 'a≠b' I have already argued that 'a+b' does not suffice. But, one should keep in mind that Smith's and Armstrong's problematic notions of necessitation permit them to speak of the existence of a *and* of b necessitating the truth of 'a≠b' without '(∃x)(x=a) & (x)(∃x=b)' entailing 'a≠b'.

3. A Trivial Dispute: Maximal vs. Minimalist Theories of Truth Making

The issue about negation reflects a general question dividing those who speak of truthmakers. As far as I know, the notion, expressed in English, was introduced by Russell in the logical atomism lectures, where he used 'makes true' and 'making true', while Moore used 'directly proves' and Russell later used 'verifier' in 1921 in the sense of 'truthmaker'. Ignoring the issue of negative facts, given an entity, (ı ß)(Ψß), and a truth, Π, we can capture the idea, and that of what others have spoken of as an ontological ground of truth, by taking (ı ß)(Ψß) to be the truthmaker for Π in the following sense:

(TM) E!(ı ß)(Ψß) & Π is true . ⊃ : (ı ß)(Ψß) makes Π true ≡. 'E! (ı ß)(Ψß)' ⊧ Π.

We need not follow Frege's line of argument against the correspondence theory of truth (or any theory for that matter) and take such a statement to itself require a truth maker as it is simply specifying or stipulating the use of the phrase '... makes ___ true' in context. Smith defines a predicate corresponding to 'makes true' in a far more complex way, but that affects nothing that follows, as far as I can see, and (TM), by employing the standard use of '⊧', avoids his problematic use of 'necessitates'.

One who holds it to be true that every true statement is such that the existence of something makes it true, must hold that for every truth Π there is some ß such that Π and ß satisfy (TM). Moreover, unlike (TM), such a purportedly true claim must itself have a truthmaker. But what could that be? Assume there are objects, properties and facts – which accounts that are concerned with truthmakers and ontologial grounds of truth accept in one form or another. Let the notion of a fact be taken broadly enough to allow for general facts or , at least, facts that serve as truth makers for true generalities. If there is a truth maker for such a maximalist claim then it is surely a fact, and not an object, property or relation. Thus the claim

(T) For every Π, Π is true iff there is a ø such that ø makes Π true

would be true in virtue of a fact, presumably a general fact. Moreover, 'is true' in (T) is either taken to stand for a property or is a predicate defined in such a way that (T) is not merely true by definition. Introducing a property of truth raises a special problem, but before turning to that there is a familiar problem to note, given the problematic status of any purported univocal truth predicate. I rule out as irrelevant the various gimmick-patterns that were generated by Kripke's attempt to invoke truth value gaps to avoid the liar paradox and yet have a univocal truth predicate in a schema. These include shifting evaluations at different assignment levels, Gupta and Belnap's notion that 'true' can be given an unproblematic yet "circular" definition, attempts to modify Tarski's conditions of satisfaction, and so on. Equally we can dismiss trivial cases, taking some truths to be true by the meaning of words, as irrelevant to the question of truthmakers. For the latter term is only viably used in an ontological sense – it is always the existence of something that makes for an ontological ground of truth.

Returning to (T), and putting aside Liar-problems, there is the obvious problem that (T) would be a truth that cannot have a truthmaker, if truth is a property, without reducing

the view to a triviality. What is required here is a general fact involving the property of *truth* and a relation of *truthmaking*. But suppose that 'is true' represents a property of truth and that there are entities (sentences – types or tokens, propositions, *etc.*) that exemplify such a property. Then if (T) is true, consider an instantiation (taking sentences, in some sense, as truth bearers, for purposes of illustration) to the name of the sentence 'Fa' – 's'. Then s, if true, exemplifies the property of truth, and it will have to be the case that its exemplifying the property of truth – i. e. there being a fact of the sentence exemplifying the property – exists only if there is something that is a truthmaker for the sentence. This means that there will be two facts relevant to 's is true'; the fact of a exemplifying F and the fact that the sentence s exemplifies the truth property. Moreover, given that s exists, it will follow that there is the one fact if and only if there is the other, for we are assuming that truth is a geunuine property that is exemplified by some kind of object and its being so forms a fact. The fact that s exemplifies *truth* will necessitate, in some sense, the fact that a is F. But this leads to the conclusion that such a theory of truthmakers reduces to the trivialization typical of so-called deflationary theories of truth that Ramsey, in his way, innaugurated. For *a* truthmaker for s would be the fact that s exemplifies truth. It is also oddly reminiscent of a Fregean account whereby 'Fa' is true in that 'Fa' denotes The True.

An odd variant of the view, taking truth as a property, is problematic for minimalist theories of truth making. I could speak of a "non-maximalist" theory, since I take the minimalist theory to simply deny (T), but take some truths to have truthmakers. The term 'minimalist' will do though, since so-called "deflationist" theories are ruled out to start with. I am concerned only with accounts that recognize things, properties and facts as grounds of truth, or constituents of such grounds, in at least some cases. Thus the "minimalist" (or "non-maximalist" view) simply involves the claim, in place of (T):

(NT) There is a Π that is true & no ø is such that ø makes Π true.

But if truth is a genuine property there will be a truthmaker for every truth – the fact that the true sentence (judgment) exemplifies the truth property. Thus given a truth property such a minimalist account is trivially inconsistent, just as a maximalist account is utterly trivial.

Suppose now there is not a property of truth – that predicates like 'is true' and 'is a truthmaker of' are not linked to properties and relations. On the maximalist account, (T) will still have a truth maker. This is prima facie problematic – that (T) is true in virtue of the existence of some fact (it would have to be a fact, if it is anything at all) without properties and relations like *being true* and *makes true*, as any attempt to specify just what such a fact would consist of without such properties (relations) seems clearly hopeless. But there is a similar problem raised by (NT). Let *s* be any truth, other than (NT) itself, that lacks a truthmaker. Then, since (NT) is an *existential* claim, on the accounts of those who speak of truth makers and the existence of something *necessitating* a truth being true, it has to be the existence of something that satisfies the existential claim of (NT). But it cannot simply be the fact that *s* exists (or *s* itself). For, to satisfy (NT), it must be its existing as lacking a truth maker. One may say that b, not the fact that b exists, is a truthmaker for the proposition 'b exists', but it has to be the fact that b is F that plays such a

truthmaking role for 'Fb', and not simply b. Likewise, it cannot be *s*'s existing or simply *s* that plays such a role in the case of (NT). Its existence will not *necessitate*, as that notion is used, (NT). That forces the proponent of (NT) to either recognize the fact that *s* is a truth without a truthmaker or to take the *sentence* (proposition, *etc.*) stating that *it is true that s is a truth without a truthmaker* to be the truthmaker for (NT). The first alternative not only introduces such a fact but also must contrue truth as a constituent property of it. The second alternative either introduces such a peculiar fact about a sentence (proposition, *etc.*) being a truth without a truthmaker – or sets off an infinite regress that never arrives at something whose existence necessitates (NT). For obviously the *mere existence* of a sentence (proposition, judgment):

'*s* is a truth without a truthmaker' is true

cannot necessitate (NT). One obviously gets nowhere by continuing to construct sentences iterating 'is true', and is thus forced, sooner or later, to acknowledge the existence of something *as a* truth without a truthmaker – i. e. a *fact* about *s*. And this means that the proponent of (NT) has to take *truth* or *makes true* or both as properties (relations) or introduce some other *fact* about *s* being true to necessitate (NT). (Necessitation is, indeed, a *real tie* .) But, as a fact about *s* being true is a truthmaker for (NT), whatever that fact is, it will satisfy the condition for being a truthmaker for *s* itself. For as it will have to involve *s existing as* a truth without a truthmaker to necessitate (NT), it will necessitate *s* (and *s*'s being true) as well. Thus (NT) is incoherent.

To suppose that *the* truth that does not have a truthmaker is (NT) itself, which, by hypothesis does not furnish the truthmaker for any truth, or to stipulate that truths without truthmakers cannot be truthmakers in turn, even if coherent, reduces the minimalist view to a triviality, as the maximalist theory was so reduced, and for a similar reason.[1]

Literature

Armstrong, D. M. 1997 *A World of States of Affairs*, Cambridge: Cambridge University Press.

Armstrong, D. M. 2000 "Difficult Cases in the Theory of Truth Makers," *The Monist*, 83, 1, 155–159.

Bergmann, G. 1992 *New Foundations of Ontology*, Madison: University of Wisconsin Press.

Carnap, R. 1947 *Meaning and Necessity*, Chicago: University of Chicago Press.

Hochberg, H. 1957 "A Note on the Empty Universe", *Mind*, 66, 264, 544–546.

Hochberg, H. 1992 "Truth Makers, Truth Predicates, and Truth Types", in K. Mulligan. (ed.) *Language, Truth and Ontology*, Dordrecht: Kluwer, 138–169.

Hochberg, H. 1999 *Complexes and Consciousness*, Stockholm: Thales.

1. I am obliged to discussion with Armstrong, Mulligan and Smith and to critical comments by Smith on an earlier version.

Mulligan, K., Simons, P. and Smith, B. 1984 "Truth-Makers", *Philosophy and Phenomenological Research*, 44, 287–321.

Owen, G. E. L. 1986 "Plato on Not-Being", in G. E. L. Owen, *Logic, Science and Dialectic: Collected Papers in Greek Philosophy*, Ithaca: Cornell University Press.

Russell, B. A. W. 1956 "The Philosophy of Logical Atomism," in B. A. W. Russell, *Logic and Knowledge*, R. Marsh (ed.) London: Allen and Unwin, 175–281.

Russell, B. A. W. 1956a "On Propositions: What They Are and How They Mean," in *Logic and Knowledge*, 285–320.

Russell, B. A. W. 1984 *Theory of Knowledge: The Collected Papers of Bertrand Russell*, v. 7, E. Eames et. al. (eds.). London: Allen & Unwin.

Searle, J. 1983 *Intentionality*, Cambridge: Cambridge University Press.

Segelberg, I. 1999 *Three Essays in Phenomenology and Ontology*, Stockholm: Thales.

Simons, P. 1992 "Logical Atomism and its Ontological Refinement: A Defense" in K. Mulligan. (ed.) *Language, Truth and Ontology*, Dordrecht: Kluwer.

Smith, B. 1999 "Truthmaker Realism", *Australasian Journal of Philosophy*, 77 (3), 274–291.

Taylor, A. E. 1971 *Plato: The Sophist and The Statesman*, R. Klibansky and E. Anscombe (eds.), London: Dawsons.

Whitehead, A. N. and Russell, B. A. W. 1950 *Principia Mathematica*, v. 1, Cambridge: Cambridge University Press.

The Two-Envelope Paradox and the Foundations of Rational Decision Theory

TERRY HORGAN

You are given a choice between two envelopes. You are told, reliably, that each envelope has some money in it – some whole number of dollars, say – and that one envelope contains twice as much money as the other. You don't know which has the higher amount and which has the lower. You choose one, but are given the opportunity to switch to the other. Here is an argument that it is rationally preferable to switch: Let x be the quantity of money in your chosen envelope. Then the quantity in the other is either 1/2x or 2x, and these possibilities are equally likely. So the expected utility of switching is 1/2(1/2x) + 1/2(2x) = 1.25x, whereas that for sticking is only x. So it is rationally preferable to switch.

There is clearly something wrong with this argument. For one thing, it is obvious that neither choice is rationally preferable to the other: it's a tossup. For another, if you switched on the basis of this reasoning, then the same argument could immediately be given for switching back; and so on, indefinitely. For another, there is a parallel argument for the rational preferability of sticking, in terms of the quantity y in the other envelope. But the problem is to provide an adequate account of how the argument goes wrong. This is the two-envelope paradox.

In an earlier paper (Horgan 2000) I offered a diagnosis of the paradox. I argued that the flaw in the argument is considerably more subtle and interesting than is usually believed, and that an adequate diagnosis reveals important morals about both probability and the foundations of decision theory. One moral is that there is a kind of expected utility, not previously noticed as far as I know, that I call *nonstandard* expected utility. I proposed a general normative principle governing the proper application of nonstandard expected utility in rational decisionmaking. But this principle is inadequate in several respects, some of which I acknowledged in a note added in press and some of which I have meanwhile discovered. The present paper undertakes the task of formulating a more adequate general normative principle for nonstandard expected utility. After preliminary remarks in section 1, and a summary in section 2 of the principal claims and ideas in Horgan 2000, I take up the business at hand in sections 3-6.

1. Preliminaries

To begin with, the paradoxical argument is an expected-utility argument. In decision theory, the notion of expected utility is commonly articulated in something like the following way (e.g., Jeffrey 1983). Let acts $A_1,\ldots,A_m$ be open to the agent, and let the agent know this. Let states $S_1,\ldots,S_n$ be mutually exclusive and jointly exhaustive possible states

B. Brogaard, B. Smith (Hg.), *Rationality and Irrationality / Rationalität und Irrationalität*, S. 172–191.

of the world, and let the agent know this. For each act A_i and each state S_j, let the agent know that if A_i were performed and S_j obtained, then the outcome would be O_{ij} and let the agent assign to each outcome O_{ij} a desirability DO_{ij}. These conditions define a *matrix formulation* of a decision problem. If the states are independent of the acts – probabilistically, counterfactually, and causally – then the *expected utility* of each act A_i is this:

$$U(A_i) = \sum_j pr(S_j) \times DO_{ij}$$

I.e., the expected utility of A_i is the weighted sum of the desirablities of the respective possible outcomes of A_i, as weighted by the probabilities of the respective possible states $S_1,\ldots,S_n$.

Second, the conditions characterizing a matrix formulation of a decision problem are apparently satisfied in the two-envelope situation, in such a way that the paradoxical argument results by applying the definition of expected utility to the relevant matrix. The states are characterized in terms of x, the quantity (whatever it is) in the agent's chosen envelope. Letting the chosen envelope be M (for 'mine') and the non-chosen one be O (for 'other'), we have two possible states of nature, two available acts, and outcomes for each act under each state, expressible this way:

	O contains 1/2x	O contains 2x
Stick	Get x	Get x
Switch	Get 1/2x	Get 2x

Matrix 1

Each of the two states of nature evidently has probability 1/2. So, letting the desirability of the respective outcomes be identical to their numerical values, we can plug into our definition of expected utility:

$$\begin{aligned} U(\text{Stick}) &= [pr(\text{O contains } 1/2x) \times D(\text{Get x})] + [pr(\text{O contains } 2x) \times D(\text{Get x})] \\ &= 1/2 \times D(\text{Get x}) + 1/2 \times D(\text{Get x}) \\ &= 1/2x + 1/2x \\ &= x \end{aligned}$$

$$\begin{aligned} U(\text{Switch}) &= [pr(\text{O contains } 1/2x) \times D(\text{Get } 1/2x)] + [pr(\text{O contains } 2x) \times D(\text{Get } 2x)] \\ &= 1/2 \times D(\text{Get } 1/2x) + 1/2 \times D(\text{Get } 2x) \\ &= 1/2 \times 1/2x + 1/2 \times 2x \\ &= 1/4x + x \\ &= 5/4x \end{aligned}$$

Third, the operative notion of probability, in the paradoxical argument and in decision theory generally, is *epistemic* in the following important sense: it is tied to the agent's total available information. So I will henceforth call it 'epistemic probability'. Although I remain neutral about the philosophically important question of the nature of epistemic

probability, lessons that emerge from the two-envelope paradox yield some important constraints on an adequate answer to that question.[1]

Fourth, below it will be useful to illustrate various points by reference to the following special case of the two-envelope decision situation, which I will call the *urn case*. Here we stipulate that the agent knows that the dollar-amounts of money in the two envelopes were determined by randomly choosing a slip of paper from an urn full of such slips; that on each slip of paper in the urn was written an ordered pair of successive numbers from the set {1,2,4,8,16,32}; that there was an equal number of slips in the urn containing each of these ordered pairs; and that the first number on the randomly chosen slip went into the envelope the agent chose and the second went into the other one. Under these conditions, the acts, states, and outcomes are represented by the following matrix:

	Stick	Switch
M contains 1 and O contains 2	Get 1	Get 2
M contains 2 and O contains 1	Get 2	Get 1
M contains 2 and O contains 4	Get 2	Get 4
M contains 4 and O contains 2	Get 4	Get 2
M contains 4 and O contains 8	Get 4	Get 8
M contains 8 and O contains 4	Get 8	Get 4
M contains 8 and O contains 16	Get 8	Get 16
M contains 16 and O contains 8	Get 16	Get 8
M contains 16 and O contains 32	Get 16	Get 32
M contains 32 and O contains 16	Get 32	Get 16

Matrix 2

Since each of the 10 state-specifications in Matrix 2 has epistemic probability 1/10,

U(Stick) = 1/10(1 + 2 + 2 + 4 + 4 + 8 + 8 + 16 + 16 + 32) = 9.3
U(Switch) = 1/10(2 + 1 + 4 + 2 + 8 + 4 + 16 + 8 + 32 + 16) = 9.3

Fifth, below I will occasionally refer to the following variant of the original two-envelope decision situation. You are given an envelope M, and there is another envelope O in front of you. You are reliably informed that M has a whole-dollar amount of money in it that was chosen by a random process; that thereafter a fair coin was flipped; and that if the coin came up heads then twice the quantity in M was put into O, whereas if the coin came up tails then half the quantity in M was put into O. I will call this the

1. Epistemic probability, as understood here, must conform to the axioms of probability theory. Although the term 'epistemic probability' has sometimes been used for subjective degrees of belief that can collectively fail to conform to these axioms, I think it is important to reclaim the term from those who have employed it that way. I would maintain that there are *objective* facts about the kind of probability that is tied to the agent's available information – i.e., about what I am calling *epistemic* probability. One objective fact is that epistemic probability obeys the axioms of probability theory.

coin-flipping situation, in contrast to the *original situation* that generates the two envelope paradox. In this coin-flipping situation, you ought rationally to switch – as has been correctly observed by those who have discussed it (e.g., Cargile 1992, 212–13, Jackson *et. al.* 1994, 44–45, and McGrew *et. al.* 1997, 29).

Finally, it also will be useful to have before us the following special case of the coin-flipping situation, which I will call the *coin-flipping urn case*. Here we stipulate that the agent knows that the whole-dollar amount in his own envelope M was determined by randomly choosing a slip of paper from an urn full of such slips; that on each slip in the urn was written one of the numbers in the set {2,4,8,16,32}; and that there was an equal number of slips in the urn containing each of these numbers. The agent also knows that after the quantity in M was thus determined, the quantity in O was then determined a fair coin-flip, with twice the quantity in M going into O if the coin turned up heads, and half the quantity in M going into O if the coin turned up tails. Under these conditions, the expected utilities are calculated on the basis of the following matrix:

	Stick	Switch
M contains 2 and O contains 1	Get 2	Get 1
M contains 2 and O contains 4	Get 2	Get 4
M contains 4 and O contains 2	Get 4	Get 2
M contains 4 and O contains 8	Get 4	Get 8
M contains 8 and O contains 4	Get 8	Get 4
M contains 8 and O contains 16	Get 8	Get 16
M contains 16 and O contains 8	Get 16	Get 8
M contains 16 and O contains 32	Get 16	Get 32
M contains 32 and O contains 16	Get 32	Get 16
M contains 32 and O contains 64	Get 32	Get 64

Matrix 3

Since the probability is 1/10 for each of the states in Matrix 3, the expected utilities are

(Stick) = 1/10(2 + 2 + 4 + 4 + 8 + 8 + 16 + 16 + 32 + 32) = 1/10(124) = 12.4
(Switch) = 1/10(1 + 4 + 2 + 8 + 4 + 16 + 8 + 32 + 16 + 64) = 1/10(155) = 15.5

2. Diagnosis and Theoretical Implications

Discussions of the two-envelope paradox (e.g., Nalebuff 1989, Cargile 1992, Castell and Batens 1994, Jackson *et. al.* 1994, Broome 1995, Arntzenius and McCarthy 1997, Scott and Scott 1997, Chalmers unpublished) typically claim that there is something wrong with the probability assignments in the paradoxical argument – although there are differences of opinion about exactly how the probabilities are supposed to be mistaken. I disagree. Consider the urn case, for example. On my construal of the paradoxical reasoning, the symbol 'x' goes proxy for a rigid definite description, which we can render as 'the ac-

tual quantity in M′ (where 'actual' is a construed as a rigidifying operator). With respect to the urn case, the following list of statements constitutes a fine-grained specification – expressed in terms of the rigid singular term 'the actual quantity in M′ – of the epistemic possibilities concerning the contents of envelopes M and O:

1. The actual quantity in M = 1 & O contains 2
2. The actual quantity in M = 2 & O contains 1
3. The actual quantity in M = 2 & O contains 4
4. The actual quantity in M = 4 & O contains 2
5. The actual quantity in M = 4 & O contains 8
6. The actual quantity in M = 8 & O contains 4
7. The actual quantity in M = 8 & O contains 16
8. The actual quantity in M = 16 & O contains 8
9. The actual quantity in M = 16 & O contains 32
10. The actual quantity in M = 32 & O contains 16

Each statement on this list has epistemic probability 1/10. Hence, since all the statements are probabilistically independent of one another, the disjunction of the five even-numbered statements on the list has probability 1/2, and the disjunction of the five odd-numbered ones also has probability one half. But the epistemic probability of the statement

O contains 1/2(the actual quantity in M)

is just the epistemic probability of the disjunction of the even-numbered statements on the list, since each even-numbered disjunct specifies one of the epistemically possible ways that this statement could be true. Likewise, the epistemic probability of the statement

O contains 2(the actual quantity in M)

is just the epistemic probability of the disjunction of the *odd*-numbered statements on the list, since each of the odd-numbered statements specifies one of the epistemically possible ways that *this* statement could be true. Therefore, in the urn case, the statements

pr(O contains 1/2(the actual quantity in M)) = 1/2
pr(O contains 2(the actual quantity in M)) = 1/2

are true. In both, the constituent statement within the scope of 'pr' expresses a *coarse-grained epistemic possibility*, a possibility subsuming exactly half of the ten equally probable fine-grained epistemic possibilities corresponding to the statements on the above list. Each of these two coarse-grained epistemic possibilities does indeed have probability 1/2, since each possibility is just the disjunction of half of the ten equally probable fine-grained epistemic possibilities. Moreover, these points about the urn case generalize straightforwardly to the original two-envelope situation. So, since the symbol 'x' in the paradoxical argument goes proxy for 'the actual quantity in M', the probability

assignments employed in the argument are correct.

How then *does* the paradoxical argument go wrong? To come to grips with this question, we need to appreciate several crucial facts about epistemic probability and about the concept of expected utility – facts that the argument helps bring into focus.

First, epistemic probability is *intensional*, in the sense that the sentential contexts created by the epistemic-probability operator do not permit unrestricted substitution *salva veritate* of co-referring singular terms. Consider the urn case, for example, and suppose that (unbeknownst to the agent, of course) the actual quantity in M is 16. Then the first of the following two statements is true and the second is false, even though the second is obtained from the first by substitution of a co-referring singular term:

pr(M contains the actual quantity in M) = 1
pr(M contains 16) = 1.

Likewise, the first of the following two statements is true and the second false, even though the second is obtained from the first by substitution of a co-referring singular term:

pr(O contains 1/2(the actual quantity in M)) = 1/2
pr(O contains 8) = ½.

It should not be terribly surprising, upon reflection, that epistemic probability is intensional in the way belief is, since epistemic probability is tied to available information in much the same way as is rational belief. (This certainly should not be surprising to those who think that epistemic probability is just rational degree of belief.)

Second, it is important to distinguish between two ways of specifying states, outcomes, and desirabilities in matrix formulations of decision problems. On one hand are *canonical* specifications: the items, as so specified, are epistemically determinate for the agent, given the total available information – i.e., the agent knows what item the specification refers to. On the other hand are *noncanonical* specifications of states, outcomes, and desirabilities: the items, as so specified, are epistemically indeterminate for the agent. The paradoxical two-envelope argument employs noncanonical specifications of states and of outcomes/desirabilities; for, the specifications employ the symbol 'x' which goes proxy for the noncanonical referring expression 'the actual quantity in M', and the quantity referred to is epistemically indeterminate (as so specified) for the agent. (The canonical/noncanonical distinction is discussed at greater length in Horgan 2000.)

Third, it needs to be recognized that because expected utility involves epistemic probabilities, and because epistemic-probability contexts are intensional, the available acts in a given decision problem can have several different kinds of expected utility. On one hand is *standard* expected utility, calculated by applying the definition of expected utility to a matrix employing canonical specifications of states, outcomes, and desirabilities. On the other hand are various kinds of *nonstandard* expected utility, calculated by applying the definition to matrices involving various kinds of noncanonical specifications.

Take the urn version of the two-envelope problem, for instance, and suppose that (unbeknownst to the agent, of course) M contains 16 and O contains 32. The standard ex-

pected utilities, for sticking and for switching, are calculated on the basis of a matrix employing canonical state-specifications, like Matrix 2 (in section 1). As mentioned above, since each of the 10 state-specifications in Matrix 2 has epistemic probability 1/10,

$$U(\text{Stick}) = 1/10(1 + 2 + 2 + 4 + 4 + 8 + 8 + 16 + 16 + 32) = 9.3$$
$$U(\text{Switch}) = 1/10(2 + 1 + 4 + 2 + 8 + 4 + 16 + 8 + 32 + 16) = 9.3$$

On the other hand, one nonstandard kind of expected utility for the acts of sticking and switching, which I will call *x-based* nonstandard utility and I will denote by 'U^x', is calculated by letting 'x' go proxy for 'the actual quantity in M' and then applying the definition of expected utility to a matrix with noncanonical state-specifications formulated in terms of x, viz., Matrix 1 (in section 1). Since each of the two state-specifications in Matrix 1 has epistemic probability 1/2, and since (unbeknownst to the agent) M contains 16,

$$U^x(\text{Stick}) = x = 16$$
$$U^x(\text{Switch}) = 1.25x = 20$$

Another nonstandard kind of expected utility for the acts of sticking and switching, which I will call *y-based* nonstandard utility and I will denote by 'U^y', is calculated by letting 'y' go proxy 'the actual quantity in O' and then applying the definition of expected utility to a matrix with noncanonical state-specifications formulated in terms of y, viz.,

	M contains 1/2y	M contains 2y
Stick	Get 1/2y	Get 2y
Switch	Get y	Get y

Matrix 4

Since each of the two state-specifications in Matrix 4 has epistemic probability 1/2, and since (unbeknownst to the agent) O contains 32,

$$U^y(\text{Stick}) = 1.25y = 40$$
$$U^y(\text{Switch}) = y = 32$$

There is nothing contradictory about these various incompatible expected-utility values for sticking and switching in this decision problem, since they involve three different kinds of expected utility – the standard kind U, and the two nonstandard kinds U^x and U^y.

Fourth, since a distinction has emerged between standard expected utility and various types of nonstandard expected utility, it now becomes crucial to give a new, more specific, articulation of the basic normative principle in decision theory – the principle of *expected-utility maximization*, prescribing the selection of an action with maximum expected utility. This principle needs to be understood as asserting that rationality requires choosing an action with maximum *standard* expected utility. Properly interpreted, therefore, the expected-utility maximization principle says nothing whatever about the various kinds of nonstandard expected utility that an agent's available acts might also happen to possess.

Having extracted these important morals about epistemic probability and expected utility from consideration of the paradoxical argument, we are now in a position to diagnose how the argument goes wrong. Since the kind of expected utility to which the argument appeals is U^x – i.e., x-based nonstandard expected utility – the principal flaw in the argument is its implicit reliance on a mistaken normative assumption, viz., that in the two-envelope decision problem, rationality requires U^x-maximization. Thus, given that U^x is the operative notion of expected utility in the paradoxical argument, the reasoning is actually correct up through the penultimate conclusion that the expected utilities of sticking and switching, respectively, are x and 1.25x. But the mistake is to infer from this that one ought to switch.

Equivocation is surely at work too. Since the unvarnished expression 'the expected utility' is employed throughout, the paradoxical argument effectively trades on the presumption that the kind of expected utility being described is *standard* expected utility. This presumption makes it appear that the normative basis for the final conclusion is just the usual principle that one ought rationally to perform an action with maximal expected utility. But since that principle applies to standard expected utility, whereas the argument is really employing a nonstandard kind, the argument effectively equivocates on the expression 'the expected utility'.

In light of this diagnosis of the paradoxical argument, and the distinction that has emerged between standard expected utility and various kinds of nonstandard expected utility, important new questions emerge for the foundations of rational decision theory: Is it sometimes normatively appropriate to require the maximization of certain kinds of nonstandard expected utility? If so, then under what circumstances?

Such questions are of interest for at least two reasons. First, the use of an appropriate kind of nonstandard expected utility sometimes provides a suitable shortcut-method for deciding on a rationally appropriate action in a given decision situation. A correctly applicable kind of nonstandard expected utility typically employs a much more coarse-grained set of states, thereby simplifying calculation.

Second (and more important), maximizing a certain kind of nonstandard expected utility sometimes is rationally appropriate in a given decision situation even though the available acts *lack* standard expected utilities – i.e., even though the total available information does not determine a uniquely rationally eligible standard probability distribution over a suitable set of exclusive and exhaustive states of the world. (By a *standard* distribution of epistemic probabilities, I mean a probability distribution over states *as canonically specified*.) In such decision situations it is rationally appropriate to maximize a certain kind of nonstandard expected utility even though the agent's total available information makes it rationally *inappropriate* to adopt any standard probability distribution, because numerous candidate-distributions all are equally rationally eligible.[2]

2. According to some construals of epistemic probability, rationality permits the initial adoption of virtually any standard probability distribution that obeys the axioms of probability and also is consistent with the agent's total available information – provided that that one then updates one's prior standard probabilities, on the basis of new evidence, in accordance with Bayes' theorem. In my view this tolerant attitude toward prior standard probabilities is mistaken, precisely because rationality prohibits the adoption of any single specific standard probability distribution when numerous candidate-distributions are all equally eligible. But even those who take

The two-envelope situation itself is a case in point. The official description of the situation does not provide enough information to uniquely fix a standard probability distribution that generates standard expected utilities for sticking and switching. (In this respect, the original problem differs from our special case, the urn version.) Nevertheless, the following is a perfectly sound expected-utility argument for the conclusion that sticking and switching are rationally on a par. Let z be the lower of the two quantities in the envelopes, so that 2z is the higher of the two. Then the epistemic possibilities for states and outcomes are described by the following matrix:

	M contains z and O contains 2z	M contains 2z and O contains z
Stick	Get z	Get 2z
Switch	Get 2z	Get z

Matrix 5

The two state-specifications in Matrix 5 both have probability 1/2. Hence the expected utility of sticking is $1/2z + 1/2(2z) = 3/2z$, whereas the expected utility of switching is $1/2(2z) + 1/2z = 3/2z$. So, since these two acts have the same expected utility, they are rationally on a par.

The soundness of this argument is commonly acknowledged in the literature. What is not commonly acknowledged or noticed, however, is that the notion of expected utility employed here is a nonstandard kind. I will call it *z-based* nonstandard expected utility, and I will denote it by U^z. In order to illustrate the fact that U^z differs from standard expected utility U, return to the urn case, and suppose that (unbeknownst to the agent, of course) the actual lower quantity z in the envelopes is 16 (and hence the actual higher quantity in the envelopes, 2z, is 32). Then, as calculated on the basis of Matrix 2 (in section 1), U(Stick) = U(Switch) = 9.3. However,

$U^z(\text{Stick}) = 1/2z + 1/2(2z) = 1/2 \times 16 + 1/2 \times 32 = 24$
$U^z(\text{Switch}) = 1/2(2z) + 1/2z = 1/2 \times 32 + 1/2 \times 16 = 24$

And with respect to the original two-envelope situation (as opposed to the urn case), there are *no* such quantities as U(Stick) and U(Switch), since there is not any single, uniquely correct, distribution of standard probabilities over canonically-specified epistemic possibilities.

The coin-flipping version of the original decision problem also illustrates the rational applicability of a suitable kind of nonstandard expected utility – in this case, E^x. The form of reasoning employed in the original two-envelope paradox not only yields the correct conclusion, but in *this* situation also appears to be a perfectly legitimate way to reason one's way to that conclusion. This fact is acknowledged in the literature; but once again, it is not commonly acknowledged or noticed that the notion of expected utility employed

the tolerant approach to prior standard probabilities can agree about the theoretical importance of nonstandard expected utilities, vis-à-vis decision situations in which the comparative rational worth of the available acts is independent of any specific standard probability distribution over epistemically determinate states of nature.

here is nonstandard. In order to illustrate the fact that U^x does indeed differ from standard expected utility, consider the coin-flipping urn case, and suppose that (unbeknownst to the agent, of course) the actual quantity in M is 16. Then although U(Stick) = 12.4 and U(Switch) = 15.5, as explained in section 1 above (with reference to Matrix 3), the x-based nonstandard expected utilities are:

$U^x(\text{Stick}) = x = 16$
$U^x(\text{Switch}) = 1/2(1/2x) + 1/2(2x) = 1/2 \times 8 + 1/2 \times 32 = 20$

And with respect to the original coin-flipping two-envelope situation (as opposed to our urn version of it), there are *no* such quantities as U(Stick) and U(Switch), since there is not any single, uniquely correct, distribution of standard probabilities over canonically-specified epistemic possibilities.

In light of these observations, in Horgan 2000 I proposed the following general normative principle for the maximization of various kinds of nonstandard expected utility in various decision situations. For a given decision problem, let δ be a singular referring expression that is epistemically indeterminate given the total available information, and hence is noncanonical. Let U^δ be a form of nonstandard expected utility, applicable to the available acts in the decision situation, that is calculated on the basis of a matrix employing noncanonical state-specifications, outcome-specifications, and desirablity-specifications formulated in terms of d. Suppose that for the given decision situation, the following *existence condition* obtains:

(E.C.) There is at least one rationally eligible standard probability distribution over epistemically possible states of nature.

Under these circumstances,

(A) Rationality requires choosing an act that maximizes U^δ just in case there is a unique ratio-scale ordering O of available acts such that (i) for every rationally eligible standard probability distribution *D* to epistemically possible states of nature for the given decision situation, U_D ranks the available acts according to O, and (ii) U^δ ranks the acts in an epistemically determinate way, and according to O.[3]

Here, U_D is the standard expected utility as calculated on the basis of *D*. To say that several standard probability assignments are "rationally eligible" does not mean, of course, that each of them is one that the agent is rationally permitted to adopt; rather, essentially it means that none of them conflict with the total available information. Insofar as they are all equally rationally eligible, it would be rationally inappropriate to adopt any one of

3. This formulation improves upon the version in Horgan 2000 by explicitly building into clause (ii) a feature that the earlier version effectively took for granted, but should have articulated: viz., that the ratio-scale ranking of available acts generated by U^δ is *epistemically determinate* for the agent (even though the U^δ-quantities themselves are epistemically indeterminate).

them, over against the others.

The proposed normative principle (A) dictates U^z-maximization in the original two-envelope situation, but not U^x-maximization or U^y-maximization. It dictates U^x-maximization in the coin-flipping version of the two-envelope situation, but not U^y-maximization or U^z-maximization. It has applications not only as an occasional short-cut method for rational decision-making that is simpler than calculating standard expected utility, but also (and much more importantly) as a method for rational decision-making in certain situations where the available acts have no standard expected utilities at all.

3. A Residual Theoretical Issue

I now think that the proposed principle (A) is inadequate, in four specific ways. I will explain the first problem in this section, and the second in section 4. In section 5 I will propose a new normative principle in place of (A), one that overcomes these two problems. Then in section 6 I will introduce the third and fourth problems, and I will address them by proposing yet another principle, a generalization of the one proposed in section 5.

The first problem is that condition (A) applies only to decision situations for which there is at least one rationally eligible standard probability distribution over epistemically possible states of nature (i.e., situations where (E.C.), the existence condition, holds). Yet there are decision situations for which (i) rationality requires choosing an act that maximizes a given kind of nonstandard expected utility, but (ii) there is no rationally eligible standard probability distribution over epistemically possible states of nature – i.e., no probability distribution over canonically specified states that satisfies all the conditions of the given decision situation. Hence, there is a need to generalize principle (A) in order to cover such decision situations.

We obtain a case in point by elaborating the original two-envelope decision situation in the following way. You are told, reliably, that the actual quantity in M has this feature: if you were to learn what it is, then you would consider it equally likely that O contains either twice that amount of half that amount; and likewise, the actual quantity in O is such that if you were to learn what it is, then you would consider it equally that M contains either twice that amount or half that amount. I will call this the *expanded version* of the two-envelope situation.

The expanded version remains a coherent decision problem. For this case too, no less that the original version, rationality requires choosing an act that maximizes U^z – which means that sticking and switching are rationally on a par. And, as in the original version, neither act has a standard expected utility. However, the reason why not is different than before. In the original version, the lack of standard expected utility was due to the fact that there were numerous rationally eligible standard probability distributions to epistemically possible states of nature – so that there is no rational reason to adopt any one of them, over against the others. In the expanded version, however, there is *no* rationally eligible standard probability distribution over the relevant states of nature. Why not? Because the following argument looms:

1. If I were to learn that M contained the minimum amount 1, then I would not consider it equally likely that O contains either twice that amount of half that amount (because I would know that O contains 2). Hence, M does not contain 1. By parallel reasoning, O does not contain 1.
2. Since neither M nor O contains 1, if I were to learn that M contained 2, then I would not consider it equally likely that O contains either twice that amount or half that amount (because I would know that O does not contain 1). Hence, M does not contain 2. By parallel reasoning, O does not contain 2.

⋮

n. Since neither M nor O contains 2n–1, if I were to learn that M contained 2n, then I would not consider it equally likely that O contains either twice that amount or half that amount (because I would know that O does not contain 2n–1). Hence, M does not contain 2n. By parallel reasoning, O does not contain 2n.

Etc.

This argument has a familiar structure: it is a version of the so-called "surprise examination paradox." Presumably it is flawed in some way – in whatever way constitutes the proper diagnosis of the surprise examination paradox. However, be that as it may, the fact that such a paradox arises from the conditions specified in the expanded two-envelope decision situation has this consequence: no standard probability distribution – i.e., no probability distribution over canonically specified potential states of envelopes M and O – is fully consistent with these specified conditions. For, the canonical state-specifications

M contains 1 and O contains 2
O contains 1 and M contains 2

each would have to be assigned probability zero, and hence the canonical state-specifications

M contains 2 and O contains 4
O contains 2 and M contains 4

each would have to be assigned probability zero, and so forth for all potential quantities in M and in O – whereas the sum of the probabilities constituting a probability distribution must be 1. So in the case of the expanded two-envelope situation, there are no rationally eligible standard probability distributions to epistemically possible states of nature.

Similar remarks apply, mutatis mutandis, to an expanded version of the coin-flipping situation that includes this additional condition: you are told, reliably, that the actual quantity in O has this feature: if you were to learn what it is, then you would consider it equally likely that M contains either twice that amount or half that amount. (Presumably it is already true, even for the earlier-described coin-flipping situation, that the actual quantity *in M* has the corresponding feature vis-à-vis O.) In this informationally enriched decision situation, as in the official coin-flipping situation, rationality requires choosing the act that maximizes U^x, viz., switching. But once again there is no rationally eligible standard probability distribution over canonically specified potential states of nature, be-

cause the conditions of the decision situation collectively have a surprise-examination structure.

I will not propose a solution to the surprise-examination paradox, nor is it one required for present purposes. The crucial points are these. First, the expanded versions of the original two-envelope situation and the coin-flipping situation are coherent decision problems, despite the fact that they have a surprise-examination structure that precludes any rationally eligible standard probability distribution over canonically-specified potential states of nature. Second, in each of these situations, rationality requires choosing an act that maximizes a certain kind of nonstandard expected utility – viz., U^z in the expanded version of the original situation, and U^x in the expanded version of the coin-flipping situation. Third, the general normative principle (A) in Horgan 2000, stating when rationality requires choosing an act that maximizes a given kind of nonstandard expected utility in a given decision situation, does not apply to the cases lately described, because principle (A) applies only when (E.C.) is satisfied – i.e., only when there is at least one rationally eligible standard probability distribution over epistemically possible states of nature. Thus arises the following theoretical issue for the foundations of rational decision theory: articulating a normative principle, to govern the application of nonstandard expected utility, that is more general than (A) – a principle that does subsume decision situations like those I have described in this section.[4]

4. A Second Residual Theoretical Issue

A second problem arises from the fact that principle (A) is supposed to specify the conditions under which rationality *requires* the maximization of a given kind of nonstandard expected utility U^δ. In a footnote to (A) in Horgan 2000, I remarked:

> Saying that rationality "requires the maximization" of U^δ means more than saying that rationality requires choosing an available act that happens to have a maximal U^δ-value. It also means that having a maximal U^δ-value is itself a *reason* why U^δ-maximization is rationally obligatory. The idea is that U^δ accurately reflects the comparative rational worth (given the agent's available information) of the available acts.

Suppose, however, that U^δ turns out to generate the right ratio-scale rankings of the actions, but for purely accidental and coincidental reasons. Then U^δ will not "accurately reflect" those rankings in the sense intended; it will not be a guaranteed, non-accidental, indicator of them. And having maximal U^δ-value will not be a "reason for rational obligatoriness," in the sense intended, to choose a U^δ-maximizing act. What is wanted, then, is something stronger that clause (ii) of principle (A). U^δ should have some feature *guaranteeing* that it generates the appropriate ranking of the available acts.

4. Note that it would not suffice merely to drop (E.C.) from the specification of the circumstances under which principle (A) applies, and leave (A) otherwise intact. For, clause (i) of principle (A) would then be *vacously* satisfied in the expanded two-envelope situation by each of U^x, U^y, and U^z. Principle (A) would thus require the maximization of *all three* of these kinds of nonstandard expected utility, in the expanded two-envelope situation – a requirement that is not only normatively inappropriate, but is impossible to fulfill.

5. Ratio-Scale Comparative Rational Worth and a New Normative Principle

One important pre-theoretic idea about rationality is that for some decision problems, the agent's total information (including desirabilities of various potential outcomes of available acts) confers upon each of the available acts some epistemically determinate, quantitively measurable, *absolute rational worth*. This idea gets explicated in decision theory in terms of the familiar notion of expected utility – i.e., what I have here called *standard* expected utility. The available acts in a given decision problem have absolute rational worth, for the agent, just in case they have standard expected utilities; and the absolute worth of each act just *is* its standard expected utility. Thus, having absolute rational worth requires that there be a set of epistemically determinate state-specifications such that (a) the agent has an epistemically determinate probability distribution over these state-specifications and (b) for each available act A_i and each state-specification, A_i has an epistemically determinate outcome and epistemically determinate desirability under the state as so specified.

Another important pre-theoretic idea about rationality is that for some decision problems, the agent's total information determines, for the set of available acts, an epistemically determinate, ratio-scale, ranking of *comparative rational worth*. When the acts each have an absolute rational worth (i.e., a standard expected utility), this will automatically confer comparative rational worth as well: the standard expected utilities fix a corresponding ratio-scale ranking of the acts. For some decision problems, however, the agent's total information determines a specific ratio-scale ranking of comparative rational worth for the available acts, *independently* of any specific probability distribution over epistemically determinate states of nature. Sometimes this happens even though there is also a uniquely correct standard probability distribution, so that the acts have standard expected utilities too (e.g., the urn case, and the coin-flipping urn case). Sometimes it happens when there is *not* a uniquely correct standard probability distribution, so that the acts do not have standard expected utilities – either (a) because the total information is consistent with more than one rationally eligible standard probability distribution over the relevant canonically specified states (e.g., the original two-envelope situation, and the coin-flipping situation), or (b) because the total available information has a "surprise examination" structure that actually *precludes* any rationally eligible probability distribution over the relevant canonically specified states (e.g., the extended versions of the original two-envelope situation and the coin-flipping situation).

Although nonstandard expected utilities are specific numerical quantities, they are epistemically indeterminate for the agent. Thus, they are not a measure of absolute rational worth. Nevertheless, in decision situations like those discussed above, nonstandard expected utility does generate an epistemically determinate ratio-scale ranking of the available acts (even though the nonstandard expected utilities themselves are epistemically indeterminate). Moreover, for each of these decision situations, the available acts stand in a unique ratio-scale ranking of comparative rational worth, independently of any specific probability distribution over canonically specified states of nature. As I will put

it, the acts stand in a unique ratio-scale ranking of *SPD-independent* comparative rational worth (i.e., comparative rational worth that is independent of any specific *standard probability distribution*). So in such decision situations, the normatively appropriate kind of nonstandard expected utility is a kind that is guaranteed to rank the available acts in accordance with their SPD-independent ratio-scale comparative rational worth. In the original two-envelope situation and its urn variant and its extended variant, U^z does this (but U^x and U^y do not), whereas in the coin-flipping situation and its urn variant and its extended variant, U^x does this (but U^y and U^z do not).

In effect, clause (i) of principle (A) is an attempt to characterize the relevant kind of SPD-independent ratio-scale comparative rational worth, and clause (ii) is an attempt to specify how a given type of nonstandard expected utility U^δ must be linked to this feature in order for U^δ-maximization to be rationally required. But clause (i) is unsatisfactory, because it fails to apply to relevant situations with a "surprise examination" structure. And clause (ii) is unsatisfactory too, because it does not preclude the possibility that U^δ happens to rank the acts in accordance with their SPD-independent ratio-scale comparative rational worth for purely fortuitous and accidental reasons. What we need, then, is a normative principle that (1) is applicable to decision situations for which there is no rationally eligible standard probability distribution over the epistemically possible states of nature (e.g., the extended two-envelope situation, and the extended coin-flipping situation), and (2) articulates the conditions under which a specific kind of nonstandard expected utility *non-accidentally* ranks the available acts in a given decision problem by SPD-independent ratio-scale comparative rational worth.

Consider the original two-envelope situation, the extended version of the original situation, and the urn case. Why is it mistaken to use U^x in these decision situations? The fundamental problem is the following. On one hand, the state-specifications employed in calculating U^x, viz.,

O contains 1/2x
O contains 2x

hold constant the epistemically indeterminate quantity x in envelope M, while allowing the content of O to vary between the two epistemically indeterminate quantities 1/2x and 2x. But on the other hand, this asymmetry, with respect to the fixity or variability of epistemically indeterminate features of the actual situation, does not reflect any corresponding asymmetry in the agent's total available information. Yet the effect of the asymmetry is that $U^x(\text{Switch}) = 5/4U^x(\text{Stick})$. Thus, since switching and sticking are rationally on a par, U^x fails to order these acts by their ratio-scale comparative rational worth.

By contrast, why is it correct to use U^z in the original two-envelope situation, in the extended version of it, and in the urn case? Because on one hand, the two state-specifications employed in calculating U^z, viz.,

M contains z and O contains 2z
M contains 2z and O contains z

are symmetric with respect to matters of fixity variability concerning the two epistemically indeterminate quantities z and 2z. The quantities themselves (viz., the lower and the higher of the two actual quantities in the two envelopes) are both held fixed; and the locations of these two quantities vary in a symmetrical way, across the two epistemically indeterminate states. On the other hand, this symmetry with respect to fixity and variability reflects the symmetry of the agent's available information concerning the contents of envelopes M and O. The result is that $U^z(\text{Switch}) = U^z(\text{Stick})$, so that U^z accurately ranks the acts in accordance with their ratio-scale comparative rational worth.

Consider now the coin-flipping version of the two-envelope situation, the extended coin-flipping version, and the coin-flipping urn case. Why is it correct to use U^x in these cases? Because on one hand, the two state-specifications employed in calculating U^x, viz.,

O contains 1/2x
O contains 2x

hold constant the epistemically indeterminate quantity x in envelope M, while allowing the content of O to vary between the two epistemically indeterminate quantities 1/2x and 2x. On the other hand, this asymmetry, with respect to the fixity and variability of epistemically indeterminate features of the actual situation, directly reflects a corresponding asymmetry in the agent's total available information: the agent knows that the quantity x in envelope M was selected first, and then either 1/2x or 2x was placed in envelope O, depending on the outcome of a fair coin-toss. That informational asymmetry renders switching 5/4 as rationally valuable as sticking. So, since the asymmetry is reflected in the fact that the state-specifications hold fixed the quantity x in envelope M while allowing the quantity in envelope O to vary between 1/x and 2x, U^x accurately ranks switching and sticking by their ratio-scale comparative rational worth: $U^x(\text{Switch}) = 5/4U^x(\text{Stick})$.

By contrast, why is it incorrect to use U^z in the coin-flipping version of the two-envelope situation, the extended coin-flipping version, and the coin-flipping urn case? Because on one hand, the two state-specifications employed in calculating U^z, viz.,

M contains z and O contains 2z
M contains 2z and O contains z

are symmetric with respect to matters of fixity and variability concerning the two epistemically indeterminate quantities z and 2z. On the other hand, these state-specifications thereby fail to reflect the crucial asymmetry in the agent's information about the contents of envelopes M and O, with the result that U^z fails to accurately rank switching and sticking by their ratio-scale comparative ratio worth of 5 to 4, and instead ranks them equally.

These observations point the way toward the general normative principle we are seeking, concerning the rational appropriateness or inappropriateness of using a specific kind of nonstandard expected utility in a given decision situation. For a given decision problem, let δ be a singular referring expression that denotes some numerical quantity and is

epistemically indeterminate given the total available information, and hence is noncanonical. Let U^{δ} be a form of nonstandard expected utility, applicable to the available acts in the decision situation, that is calculated on the basis of a matrix employing noncanonical state-specifications, outcome-specifications, and desirability-specifications formulated in terms of δ. We will say that the set of state-specifications employed to calculate U^{δ} is *symmetry and asymmetry reflecting, with respect to fixity and variability of features of the decision situation* (for short, $SAR_{f/v}$) just in case any symmetries or asymmetries in these state-specifications reflect corresponding symmetries and asymmetries in the agent's total available information. Then

(B) Rationality requires choosing an act that maximizes U^{δ} if (i) U^{δ} employs state-specifications that are $SAR_{f/v}$, and (ii) U^{δ} generates an epistemically determinate ratio-scale ranking of the available acts.[5]

When the conditions in (B) are met, the available acts do indeed possess SPD-independent ratio-scale comparative rational worth, and U^{δ} is guaranteed to rank the acts in a way that accurately reflects their comparative worth. For, the very symmetries and asymmetries in the agent's total information that fix determinate ratio-scale comparative worth for the acts, independently of any specific probabilities for canonical state-specifications, are directly reflected in the fixity/variability structure of the noncanonical state-specifications employed by U^{δ}.

6. Ordinal-Scale Rational Worth and a More General Normative Principle

Although the two problems with principle (A) described in sections 3 and 4 have now been dealt with, two further problems need to be addressed; both also arise for principle (B) and hence will prompt modifications of (B) in turn. The third problem is that there are decision problems for which (i) the available acts stand in an *ordinal-scale*, but not a *ratio-scale*, ordering of SPD-independent comparative rational worth, and (ii) there is a suitable kind of nonstandard expected utility that rationally ought to be maximized (because it is guaranteed to reflect the ordinal-scale comparative rational worth of the acts).

Here is a simple example. You are given a choice between two envelopes E1 and E2, after being reliably informed that first some whole-dollar quantity of money of $2 or more was chosen by some random process and placed in E1, and then the square of that quantity was placed into E2. Assuming that the desirability of an outcome is just the dollar-amount obtained, in this decision situation there is a kind of nonstandard expected utility definable for this situation that ought rationally to be maximized, viz. U^{w}, where w = the actual quantity in E1. Since

$U^{w}(\text{Choose E1}) = w$

5. Condition (B) is stated merely as a sufficient condition for the rationality of U^{δ}-maximization, rather than a sufficient *and necessary* condition, because it is still not general enough to cover all cases. See section 6.

$U^{w}(\text{Choose E2}) = w^2$

and since $w^2 > w$ for all potential values of w, rationality requires the U^{w}-maximizing act, viz., choosing E2. However, since the epistemically possible quantities in E2 are a nonlinear function of the corresponding epistemically possible quantities in E1, the two acts do not stand not in an SPD-independent *ratio-scale* ranking of comparative rational worth, but only in an SPD-independent *ordinal-scale* SPD-independent ranking of comparative worth. (Accordingly, U^{w} generates only an epistemically determinate ordinal-scale ranking of the acts.)

The fourth problem is that rationality sometimes requires maximizing a more general version of nonstandard expected utility than has so far been discussed, a version involving several noncanonical number-denoting terms rather than just one. Consider the following decision situation, for example. You are given a choice of two envelopes E1 and E2. Envelope E1 has two slots $S1_{E1}$ and $S2_{E1}$, and envelope E2 has two slots $S1_{E2}$ and $S2_{E2}$. Each slot in E1 contains some dollar-quantity of money, selected by some random process. (The two selections were independent of one another.) Slot $S1_{E2}$ of E2 contains either half or twice the quantity in slot $S1_{E1}$ of E1, depending on the outcome of a fair coin-flip. Slot $S2_{E2}$ of E2 contains either one fourth of, or four times, the quantity in slot $S1_{E2}$, depending on the outcome of an independent fair coin-flip.

Letting x be the actual quantity in $S1_{E1}$ and y be the actual quantity in $S2_{E1}$, there is a nonstandard expected utility $U^{x,y}$ definable for this decision problem that yields epistemically indeterminate expected utilities expressed as mathematical functions of x and y. Assuming that the desirabilities of the potential outcomes are just their dollar amounts,

$U^{x,y}(\text{Choose E1}) = 1/4[(x + y) + (x + y) + (x + y) + (x + y)] = x + y$
$U^{x,y}(\text{Choose E2}) = 1/4[(1/2x + 1/4y) + (2x + 1/4y) + (1/2x + 4y) + (2x + 4y)] = 5/4x + 17/8y.$

$U^{x,y}$ is guaranteed to reflect the acts' SPD-independent comparative ordinal-scale rational worth, because $(5/4x + 17/8y) > (x + y)$ for any permissible values of x and y. Thus, rationality dictates the maximization of $U^{x,y}$ in this decision situation. (Notice that the third problem too is illustrated by this case. The stated conditions fix an SPD-independent *ordinal-scale* comparative rational worth for the two acts, without fixing any unique ratio-scale ordering: choosing E2 is rationally preferable to choosing E1, but not by any specific, probability-independent, ratio.)

So for some decision problems, a certain kind of nonstandard expected utility reflects SPD-independent ordinal-scale comparative rational worth of the available acts, even when they lack PDP-independent *ratio-scale* comparative rational worth. Moreover, for some decision problems, SPD-independent comparative rational worth is reflected by a kind of nonstandard expected utility based on several noncanonical number-denoting terms rather than one. Thus a normative principle more general than (B) is needed, to govern the rationally appropriate use of nonstandard expected utility in such cases.

The needed principle can be articulated by generalizing (B) in the following way. For a given decision problem, let $\delta 1,\ldots,\delta m$ be singular referring expressions that denote numerical quantities and are epistemically indeterminate given the total available information, and hence are noncanonical. Let $U^{\delta 1,\ldots,\delta m}$ be a form of nonstandard expected utility,

applicable to the available acts in the decision situation, that is calculated on the basis of a matrix employing noncanonical state-specifications, outcome-specifications, and desirability-specifications formulated in terms of $\delta 1,\ldots,\delta m$. Then

(C) Rationality requires choosing an act that maximizes $U^{\delta 1,\ldots,\delta m}$ just in case (i) $U^{\delta 1,\ldots,\delta m}$ employs state-specifications that are $SAR_{f/v}$, and (ii) $U^{\delta 1,\ldots,\delta m}$ generates an epistemically determinate ordinal-scale ranking of the available acts.[6]

When these conditions are met, the available acts do indeed possess SPD-independent ordinal-scale comparative rational worth, and $U^{\delta 1,\ldots,\delta m}$ is guaranteed to rank the available acts in a way that accurately reflects their comparative rational worth. For, the very symmetries and asymmetries in the agent's total information that fix determinate ordinal-scale comparative worth for the acts, independently of any specific probabilities for canonical state-specifications, are directly reflected in the fixity/variability structure of the noncanonical state-specifications employed by $U^{\delta 1,\ldots,\delta m}$. So we have arrived at a general normative principle governing the maximization of nonstandard expected utility, a principle that overcomes all four problems faced by principle (A).

Principle (B), which states only a sufficient condition for the rationality of maximizing a given kind of nonstandard expected utility (rather than a sufficient *and necessary* condition), remains in force. In effect, it is a special case of our more general normative principle (C).

Let me make several final observations about principles (C) and (B) and the key notion they employ, viz., the feature $SAR_{f/v}$. First, I take it that the failure to be $SAR_{f/v}$ is a feature that can be exhibited only by state-specifications of the kind that figure in *nonstandard* expected utility, viz., *epistemically indeterminate* state-specifications. Only when relevant features of the actual situation are specified in epistemically indeterminate ways does it become possible to fix or vary them in ways not reflective of one's total information, within a set of state-specifications that are mutually exclusive and jointly exhaustive.

Second, the feature of being $SAR_{f/v}$ is evidently clear enough to be useful and applicable in concrete decision situations like those I have described in this paper. Often in such situations, one can tell by inspection whether or not the state-specifications employed by a given kind of nonstandard expected utility are $SAR_{f/v}$. Indeed, it is evidently very common in practice – in betting decisions, for example – to rely on calculations of nonstandard expected utilities that are $SAR_{f/v}$.

6. Clause (ii) is non-redundant, because there are decision situations in which clause (i) is satisfied but clause (ii) is not. Here is an example. You are given a choice between two envelopes E1 and E2, each of which contains some whole-dollar quantity of money. You are told that some quantity n, evenly divisible by 3, was first selected by a random process and placed into E1, and that the quantity $(n/3)^2$ was then placed into E2. Letting w = the actual quantity in E1, U^w(Choose E1) = w, whereas U^w(Choose E2) = $(w/3)^2$. In this situation U^w is a form of nonstandard expected utility that satisfies clause (i) of principle (C). However, U^w does not generate an epistemically determinate ordinal-scale ranking of the available acts, and hence does not satisfy clause (ii) of (C). For, U^w(Choose E1) > U^w(Choose E2) if $w < 9$, whereas U^w(Choose E1) = U^w(Choose E2) if $w = 9$, whereas U^w(Choose E1) < U^w(Choose E2) if $w > 9$.

But third, being $SAR_{f/v}$ also has been characterized somewhat vaguely, in terms of several vague ideas: (1) symmetries and asymmetries in one's total information, (2) symmetries and asymmetries in a set of noncanonical state-specifications, and (3) a relation of "reflection" between the latter and the former kinds of symmetries and asymmetries. It would be theoretically desirable to explicate these notions further, and to employ the explicated versions to articulate a sharpened normative principle that would replace and explicate the vague normative principles (C) and (B).

Fourth, the notion of SPD-independent comparative rational worth is also somewhat vague, as so far characterized. It would be theoretically desirable to provide a direct explication of it too, and to explicitly articulate its connection to explicated versions of principles (C) and (B). These tasks of further explication and articulation I leave for a future occasion.[7]

References

Arntzenius, F. and McCarthy, D. 1997 "The Two Envelope Paradox and Infinite Expectations," *Analysis*, 57, 42–50.

Broome, J. 1995 "The Two-Envelope Paradox," *Analysis*, 55, 6–11.

Cargile, J. 1992 "On a Problem about Probability and Decision," *Analysis*, 54, 211–16.

Castell, P. and Batens, D. 1994 "The Two-Envelope Paradox: The Infinite Case," *Analysis*, 54, 46–49.

Chalmers, D. Unpublished "The Two-Envelope Paradox: A Complete Analysis?"

Horgan, T. 2000 "The Two-Envelope Paradox, Nonstandard Expected Utility, and the Intensionality of Probability," *Nous*, 34, 578–602.

Jeffrey, R. 1983 *The Logic of Decision*, Second Edition, Chicago: University of Chicago Press.

Jackson, F., Menzies, P., and Oppy, G. 1994 "The Two Envelope 'Paradox'," *Analysis*, 54, 43–45.

McGrew, T., Shier, D. and Silverstein, H. 1997 "The Two-Envelope Paradox Resolved," *Analysis*, 57, 28–33.

Nalebuff, B. 1989 "The Other Person's Envelope is Always Greener," *Journal of Economic Perspectives*, 3, 171–81.

Scott, A. and Scott, M. 1997 "What's in the Two Envelope Paradox?" *Analysis*, 57, 34–41.

7. I dedicate this paper to my wife Dianne, who has patiently endured my envelope obsession. She plans to put my ashes into two envelopes, and then put one envelope on the mantel and sprinkle the other's contents into the wind at the U.S. Continental Divide.

The Rationality of Reasoning: Commitment and Coherence

JOHN KEARNS

Reasoning, especially correct deductive reasoning, is surely a paradigm of rationality, or rational activity, for reasoning is the apparent source of our fundamental concept of the rational. In logic, one studies correct deductive reasoning, among other things, but standard logic is poorly equipped to understand and explain this reasoning. For the norms that govern deductive reasoning depend on more than truth and truth conditions, while standard logic is fixated on truth conditions. What I understand by 'illocutionary logic' is a broader subject matter than standard logic, and possesses the resources to give an adequate account of the rationality of reasoning.

In logic one focuses on the linguistic. This isn't a problem; most deductively correct reasoning is carried out with language – and all deductively correct reasoning can be formulated in language. However, I take the fundamental linguistic reality to consist of speech acts, or linguistic acts, and not of mere expressions. Linguistic acts are the primary bearers of semantic features such as meaning, truth, and falsity. To consider issues of rationality as they relate to reasoning, I shall attend to linguistic acts rather than to languages considered abstractly.

Someone who uses an expression to perform a meaningful act has performed a linguistic act, a speech act. She can do this in speaking or writing, or in thinking words "in her head." A person who reads or who listens with understanding is also performing linguistic acts. *Sentential acts* are performed with sentences. Those sentential acts that can appropriately be evaluated in terms of truth and falsity are *propositional acts*, or *statements*. (This is a stipulative use for 'statement,' because, conversationally, 'statement' is often a near synonym of 'assertion.') Much work in logic sheds light on statements and on argument sequences whose components are statements.

A propositional act can be performed with a certain *illocutionary force*. A statement may be an assertion or a denial, it can merely be supposed. The illocutionary force of a statement is distinct from its character as a statement, for it isn't essential that a statement have illocutionary force. Someone might make a statement merely to consider it. And a statement which is one disjunct of an assertion will not itself be asserted.

An argument which moves from premisses to a conclusion whose truth they support (or purport to) is itself a linguistic act, although it is not propositional. A *simple* argument has component premisses and a conclusion which are propositional illocutionary acts – they might be assertions or denials or suppositions. A *complex* argument contains other arguments as components.

Our practices of using language for thinking, for communicating, and for making arguments, among other things, are "governed" by norms. There are correct and incorrect ways to use expressions, there are correct and incorrect assertions, and correct and incorrect arguments. One logical enterprise is that of discovering, developing, and investigating the norms governing some linguistic activities. It is not entirely felicitous to speak of logics, or to ask which is the right logic. But we can appropriately ask which logical system best describes or captures a linguistic practice. Alternative systems of logic might be suited to different linguistic practices, or to different aspects of one practice.

B. Brogaard, B. Smith (Hg.), *Rationality and Irrationality / Rationalität und Irrationalität*, S. 192–198.

A logical system, or logical theory, has three components: an artificial language, a semantic account which gives the truth conditions of sentences in the language, and a deductive system for codifying certain items of the artificial language (the logically true sentences, the logically valid argument sequences, etc.). Strictly speaking, an artificial language isn't really a language, for no one speaks it or thinks it. From a speech-act perspective, it is most appropriate to consider sentences in an artificial language as representations of statements in natural languages (as representations of *kinds* of statements). The truth conditions of these sentences are for the statements that they represent. From this perspective, a logical theory is an empirical theory designed to capture a certain human practice – but this is a normative practice, and some conceptual analysis will be necessary to determine if a given theory captures a given practice.

In the references listed at the end of this paper, I have developed systems of illocutionary logic. These differ from standard logical systems in three respects:

(1) Illocutionary force indicating expressions, or, simply, *illocutionary operators*, are added to the artificial language;
(2) The account of truth conditions, the truth-conditional semantics for the artificial language, is supplemented by an account of commitment conditions; these determine what statements a person is committed to accept or reject once she accepts and rejects some to begin with;
(3) The deductive system is modified to take account of, and accommodate, illocutionary operators.

So that we have an example to consider, let L be the artificial language of propositional logic which contains atomic sentences and connectives for negation ($\sim$), disjunction ($\vee$), and conjunction ($\&$). The horseshoe of material implication ($\supset$) is a defined symbol. The sentences of L have the customary truth conditions, characterized by familiar truth-tables. In considering which items in L to codify by means of a deductive system, we might choose to focus on sentences and on the *plain argument sequences* characterized as follows:

> If $A_1,\ldots, A_n, B$ are sentences of L, then '$A_1,\ldots, A_n / B$' is a *plain argument sequence*. The sentences $A_1,\ldots, A_n$ are the *premisses* and B is the *conclusion*.

These sequences consist of 0 or more premisses (in a given order) followed by a conclusion. Given the truth conditions of sentences of L, we can define concepts of logical truth for sentences and logical validity for plain argument sequences.

Since the results of Tarski anyway, most work in logic has focused on truth conditions and on concepts defined in terms of truth conditions. This doesn't provide an adequate analysis or explanation of the normative and rational practice of constructing deductive arguments. Someone who reasons correctly from the premisses to the conclusion of an ar-

gument is not propelled by truth or truth conditions. She must rely on her understanding of the connection between premisses and conclusion. This may well involve her understanding of truth conditions, but this "cashes out" in terms of what I call *rational commitment*. She understands that once she has accepted or supposed the premisses, she is committed to accept the conclusion or to give it the force of a supposition.

Rational commitment must be distinguished from moral or ethical commitment. Someone with a moral commitment has an obligation to perform or not perform some course of action. Two people who make a moral commitment to each other are obligated in various ways to one another. Rational commitment is generated by decisions, and by other intentional acts or activities. But it is not a moral obligation to carry out the action one is committed to perform. Someone who decides to buy gas on the way to work has established a rational commitment to do this. If she forgets, and reaches her place of employment without buying gas, she may kick herself for being forgetful, but she hasn't done anything immoral.

A rational commitment can be absolute, like the commitment to buy gas, or conditional, like the commitment to answer the phone if it rings. (Someone who is rationally committed to do *X* may also be morally obligated to do *X*, but the commitment does not generate the obligation.) A rational commitment, as I am understanding it, is always a commitment to do or not do something. People sometimes describe a person as being committed to the truth of some statement. This is not my sort of commitment. In my sense, a person might be committed to *acknowledge* the truth of a statement.

Deciding to do *X* commits a person to doing *X*. But accepting or rejecting one statement will also commit a person to accepting or rejecting others, so long as she persists in her original assertion/acceptance or denial/rejection (so long as she doesn't forget or change her mind). Supposing that some statements are true, or false, commits a person to granting other statements the status of suppositions. We don't usually call the others suppositions, but I will stretch the use of 'suppose' and 'supposition' to cover those cases. On this usage, making some suppositions will commit a person to making others. The commitment to accept/assert or reject/deny or suppose is ordinarily conditional – accepting the statement that today is Thursday commits me (today) to accepting the statement that tomorrow is Friday, but only if the matter comes up and I choose to think about it.

The truth conditions for sentences of *L* are familiar, but we also need a *commitment semantics* for *L*. Commitment, unlike truth and truth conditions, is relative to a person or community. The commitment semantics will be developed from the perspective of an idealized *designated subject*. I use '+' for the value of sentences (statements) which the designated subject is committed to accept, '–' for those she is committed to reject (as false), and '*n*' for the rest. The following matrices give a partial characterization of the commitment semantics for *L*.

A	*B*	~*A*	[*A* ∨ *B*]	[*A* & *B*]	[*A* ⊃ *B*]
+	+	–	+	+	+
+	*n*	–	+	*n*	*n*
+	–	–	+	–	–
n	+	*n*	+	*n*	+
n	*n*	*n*	+,*n*	*n*,–	+,*n*
n	–	*n*		–	*n*
–	+	+	+	–	+
–	*n*	+	*n*	–	+
–	–	+	–	–	+

As can be seen, in some cases the commitment values of components do not completely determine the values of compound sentences. The matrices must be supplemented to fully characterize a commitment valuation for *L*. Such an account is found in the first two references at the end of this paper.

We also need to add *illocutionary operators* to *L*. There are four illocutionary operators in *L*:

⊦ – for asserting/accepting a statement
⊣ – for denying/rejecting a statement
⌊ – for supposing a statement true
¬ – for supposing a statement false

The operators for assertion/acceptance and denial/rejection are Frege's. The operator for supposing true is mine, and the operator for supposing false is borrowed from Intuitionist logic. We need these operators if we are to represent realistic (speech-act) arguments. Statements occurring in arguments have illocutionary force, and criteria for deductive correctness must take account of illocutionary force.

Now that I have listed the illocutionary operators, I need to give more information about the language *L*. The sentences of *L* that were considered earlier are *plain* sentences; these are the atomic sentences and the compound sentences constructed with atomic sentences and connectives. If we prefix a plain sentence of *L* with an illocutionary operator, the result is a *completed* sentence. This is the only way to obtain completed sentences. (The illocutionary operators cannot be iterated, and completed sentences cannot be components of larger sentences.)

If we focus on plain sentences, we find a difference between truth-conditional implication and two commitment-based concepts. Definitions of these concepts are given as follows:

Sentences (statements) $A_1,\ldots, A_n$ *truth-conditionally imply* B iff it isn't possible for $A_1,\ldots, A_n$ to be true when B isn't.

Sentences $A_1,\ldots, A_n$ *basically imply* B iff accepting $A_1,\ldots, A_n$ commits a person to accept B.
Sentences $A_1,\ldots, A_n$ *suppositionally imply* B iff supposing $A_1,\ldots, A_n$ to be true commits a person to suppose B to be true.

These definitions give the basic ideas of the various concepts. The concepts can also be characterized formally in terms of truth-conditional interpreting functions and commitment valuations, but these details can be omitted for the present.

It is completed sentences that receive the most attention in illocutionary logic. To give them their due, we require semantic concepts for which there are no natural truth-conditional counterparts. For example, we will be interested in determining the consequences of a combination of assertions, denials, and positive and negative suppositions. I will use the expression 'logical requirement' for the appropriate concept. A definition can be given as follows:

Completed sentences $A_1,\ldots, A_n$ *logically require* completed sentence B iff performing the acts $A_1,\ldots, A_n$ commits a person to performing the act B.

This concept can also be formally characterized. Given this definition, we can see that $\vdash A$ and $\dashv[A \,\&\, B]$ logically require $\dashv B$.

With completed sentences, there is a need to consider a different kind of argument sequence than before, an *illocutionary (argument) sequence*:

If $A_1,\ldots, A_n$, B are completed sentences, then '$A_1,\ldots, A_n \rightarrow B$' is an *illocutionary (argument) sequence*.

An illocutionary sequence is *logically connected* iff its premisses logically require its conclusion.

In the deductive system for an illocutionary logical theory, it seems most appropriate to codify those assertions, denials, and illocutionary sequences that have a logically distinguished character.

A simple argument whose premisses truth-conditionally imply its conclusion might not be correct. The argument:

$$\frac{\llcorner A \quad \llcorner B}{\vdash[A \,\&\, B]}$$

is incorrect. Although the premisses cannot be true without the conclusion being true, the premisses are supposed, and these suppositions will not justify asserting the conclusion. Validity is not an appropriate concept (standard?) for evaluating arguments understood as speech acts. What we want from an argument with deductive intentions is *deductive correctness*: The premiss acts, with their forces, must commit the arguer to performing the conclusion act with its force; i.e. the premiss acts must logically require the conclusion act. These arguments are deductively correct:

$$\frac{\vdash A \quad \vdash B}{\vdash[A \,\&\, B]} \qquad \frac{\vdash A \quad \llcorner B}{\llcorner[A \,\&\, B]} \qquad \frac{\llcorner A \quad \llcorner B}{\llcorner[A \,\&\, B]}$$

Perceived rational commitment is the motive power that propels an arguer from premisses to conclusion. And commitment-based concepts provide the appropriate standards for evaluating deductive arguments. The reciprocal abilities to make and to recognize rational commitment are the abilities that give rise to reason and reasoning. And illocutionary logic is the subject matter which investigates commitment-based concepts as well as the relations between commitment conditions and truth conditions.

Performing some acts can rationally commit a person to perform or to not perform others. But there is a general, come what may, rational commitment to act *coherently*. Let us consider how this applies to illocutionary acts. I shall use the word 'consistent' for a semantic concept based on truth conditions. Statements that might be true together are *consistent* with one another. But propositional illocutionary acts are *coherent* or not. Accepting or supposing inconsistent statements is *incoherent*. It is also incoherent to perform acts which commit one to perform incoherent acts.

It can sometimes be incoherent to accept consistent statements. G. E. Moore was puzzled by such statements as "It's raining, but I don't believe it." They are consistent by my definition, but it makes no sense for a person to accept one of these statements. Even though the statements made in (a) and (b):

(a) $\vdash$ *It is raining.*
(b) $\vdash\sim$ *I believe that it is raining.*

are consistent with one another, the illocutionary acts are incoherent. For assertion (a) commits the speaker to performing assertion (c):

(c) $\vdash$ *I believe that it is raining.*

whose statement is inconsistent with that in assertion (b).

Supposing inconsistent statements is also incoherent. It is legitimate to do this in the course of making an indirect proof (a proof by contradiction), but the general commitment to coherence is what leads one to give up an incoherent hypothesis, and reach the opposite conclusion. Illocutionary logic is also the appropriate subject matter for exploring coherence and incoherence.

There would be no rational commitment if people were not equipped to make and to recognize this commitment. Rational commitment is the foundation and source of rationality. With respect to true or false statements, rational commitment often tracks truth conditions, as in the commitment from accepting A and accepting B to accepting the conjunction of A with B. But some commitment is not truth-preserving. The commitment from accepting a statement A to accepting '*I believe that A*,' is not truth-preserving, since there are many true statements that I don't believe. Similarly, the commitment from accepting '*I believe that A*' to accepting A is not truth-preserving. And rational commitment

Sentences $A_1,\ldots, A_n$ *basically imply* B iff accepting $A_1,\ldots, A_n$ commits a person to accept B.
Sentences $A_1,\ldots, A_n$ *suppositionally imply* B iff supposing $A_1,\ldots, A_n$ to be true commits a person to suppose B to be true.

These definitions give the basic ideas of the various concepts. The concepts can also be characterized formally in terms of truth-conditional interpreting functions and commitment valuations, but these details can be omitted for the present.

It is completed sentences that receive the most attention in illocutionary logic. To give them their due, we require semantic concepts for which there are no natural truth-conditional counterparts. For example, we will be interested in determining the consequences of a combination of assertions, denials, and positive and negative suppositions. I will use the expression 'logical requirement' for the appropriate concept. A definition can be given as follows:

Completed sentences $A_1,\ldots, A_n$ *logically require* completed sentence B iff performing the acts $A_1,\ldots, A_n$ commits a person to performing the act B.

This concept can also be formally characterized. Given this definition, we can see that $\vdash A$ and $\dashv[A \,\&\, B]$ logically require $\dashv B$.

With completed sentences, there is a need to consider a different kind of argument sequence than before, an *illocutionary (argument) sequence*:

If $A_1,\ldots, A_n$, B are completed sentences, then '$A_1,\ldots, A_n \rightarrow B$' is an *illocutionary (argument) sequence*.

An illocutionary sequence is *logically connected* iff its premisses logically require its conclusion.

In the deductive system for an illocutionary logical theory, it seems most appropriate to codify those assertions, denials, and illocutionary sequences that have a logically distinguished character.

A simple argument whose premisses truth-conditionally imply its conclusion might not be correct. The argument:

$$\frac{\llcorner A \quad \llcorner B}{\vdash[A \,\&\, B]}$$

is incorrect. Although the premisses cannot be true without the conclusion being true, the premisses are supposed, and these suppositions will not justify asserting the conclusion. Validity is not an appropriate concept (standard?) for evaluating arguments understood as speech acts. What we want from an argument with deductive intentions is *deductive correctness*: The premiss acts, with their forces, must commit the arguer to performing the conclusion act with its force; i.e. the premiss acts must logically require the conclusion act. These arguments are deductively correct:

$$\frac{\vdash A \quad \vdash B}{\vdash[A \,\&\, B]} \qquad \frac{\vdash A \quad \llcorner B}{\llcorner[A \,\&\, B]} \qquad \frac{\llcorner A \quad \llcorner B}{\llcorner[A \,\&\, B]}$$

Perceived rational commitment is the motive power that propels an arguer from premisses to conclusion. And commitment-based concepts provide the appropriate standards for evaluating deductive arguments. The reciprocal abilities to make and to recognize rational commitment are the abilities that give rise to reason and reasoning. And illocutionary logic is the subject matter which investigates commitment-based concepts as well as the relations between commitment conditions and truth conditions.

Performing some acts can rationally commit a person to perform or to not perform others. But there is a general, come what may, rational commitment to act *coherently*. Let us consider how this applies to illocutionary acts. I shall use the word 'consistent' for a semantic concept based on truth conditions. Statements that might be true together are *consistent* with one another. But propositional illocutionary acts are *coherent* or not. Accepting or supposing inconsistent statements is *incoherent*. It is also incoherent to perform acts which commit one to perform incoherent acts.

It can sometimes be incoherent to accept consistent statements. G. E. Moore was puzzled by such statements as "It's raining, but I don't believe it." They are consistent by my definition, but it makes no sense for a person to accept one of these statements. Even though the statements made in (a) and (b):

(a) $\vdash$ *It is raining.*
(b) $\vdash\sim$ *I believe that it is raining.*

are consistent with one another, the illocutionary acts are incoherent. For assertion (a) commits the speaker to performing assertion (c):

(c) $\vdash$ *I believe that it is raining.*

whose statement is inconsistent with that in assertion (b).

Supposing inconsistent statements is also incoherent. It is legitimate to do this in the course of making an indirect proof (a proof by contradiction), but the general commitment to coherence is what leads one to give up an incoherent hypothesis, and reach the opposite conclusion. Illocutionary logic is also the appropriate subject matter for exploring coherence and incoherence.

There would be no rational commitment if people were not equipped to make and to recognize this commitment. Rational commitment is the foundation and source of rationality. With respect to true or false statements, rational commitment often tracks truth conditions, as in the commitment from accepting A and accepting B to accepting the conjunction of A with B. But some commitment is not truth-preserving. The commitment from accepting a statement A to accepting '*I believe that A*,' is not truth-preserving, since there are many true statements that I don't believe. Similarly, the commitment from accepting '*I believe that A*' to accepting A is not truth-preserving. And rational commitment

links other kinds of illocutionary acts than those performed with true or false statements. (Such as questions and commands.) Illocutionary logic is not at odds with standard logic, for illocutionary logic contains standard logic as a proper part, but it is the illocutionary logic that goes beyond standard logic which investigates the foundations of rationality.

The rationality embodied in rational commitment is a matter of intelligence or good sense, but not of morality. My intuitions are that moral values, virtues, and obligations are not directly implicated by considerations of rational commitment. Although not immoral, it makes no sense to accept *A*, to realize that accepting *A* commits you to accepting *B*, and nonetheless to refuse to accept *B*. Someone who fails to "honor" rational commitment, or who acts or believes incoherently, is foolish, perhaps flawed, but not *bad*.

There are some epistemological theories that speak of obligations having a somewhat moral character. These theories may claim that there are some beliefs that one *ought* to hold. However, my own view is that any epistemic obligation can at most command us not to acquire beliefs in careless or capricious ways. Such requirements do not confer an obligatory status on honoring one's rational commitments. An obligation to acquire only justified beliefs does not determine which beliefs in particular one should end up with.

Literature

Kearns, J. T. 1977 "Propositional Logic of Supposition and Assertion," *Notre Dame Journal of Formal Logic*, 38, 325–349.

Kearns, J. T. 2000 "An Illocutionary Logical Explanation of the Surprise Execution," *History and Philosophy of Logic* 20, 195–214.

Kearns, J. T. 2000 "The Priority of Denial to Negation," presented at the *2000 Annual Meeting of the Association for Symbolic Logic*, the University of Illinois at Urbana-Champaign. An abstract will appear in *The Journal of Symbolic Logic*.

Welche Stufe der Rationalität ist in Recht und Ethik erreichbar und wünschenswert?

EDGAR MORSCHER

1. Einleitung

Mit ‚Rationalität' ist im folgenden – zumindest primär – weder die Rationalität einer Person noch die Rationalität ihrer Überzeugungen oder Handlungen gemeint. Vielmehr geht es mir hier um das Problem, ob und – falls ja – in welchem Ausmaß eine Frage rational beantwortet werden kann. Fragen, die überhaupt nicht rational behandelt und beantwortet werden können, sind reine Geschmacksfragen; und über bloße Geschmacksfragen läßt sich bekanntlich zwar trefflich streiten, aber nicht (rational) argumentieren.

Als Musterbeispiel an Rationalität gelten die Wissenschaften mit ihren Problemstellungen. Aber auch innerhalb der Wissenschaften gibt es verschiedene Stufen an Rationalität je nach Fachgebieten. In gewissen Bereichen von Logik und Mathematik können wir offene Fragen dadurch entscheiden, daß wir Antworten, die dafür vorgeschlagen werden, beweisen oder widerlegen. Der damit verbundene Rationalitätsstandard ist besonders hoch, vielleicht der höchste überhaupt, der erreicht werden kann. Er wurde früher für alle Wissenschaften und alle wissenschaftlichen Fragestellungen in Anspruch genommen: Lange galt die Beweisbarkeit und Widerlegbarkeit ihrer Sätze sogar als Markenzeichen aller Wissenschaften; heute wissen wir jedoch, daß dieses Ideal nicht einmal in Logik und Mathematik – geschweige denn in den anderen Wissenschaften – allgemein erreichbar ist. Ähnlich hielt man in den empirischen Wissenschaften die Verifizierbarkeit ihrer Sätze lange als gültigen Rationalitätsstandard; demgegenüber gilt heute die Nicht-Verifizierbarkeit, ja sogar die Unterbestimmtheit von (natur-)wissenschaftlichen Theorien durch Daten als unbestritten, und wir begnügen uns mit ihrer Konfirmierbarkeit und Diskonfirmierbarkeit. Übersetzungen bleiben nicht nur – wie die strikt universellen Hypothesen der empirischen Wissenschaften – durch Daten unterbestimmt, sondern für sie gilt – nach Quine – sogar die These der Unbestimmtheit; dennoch entziehen sie sich deswegen nicht jedem Rationalitätsanspruch. Für die normativen Sätze in Recht und Ethik gilt meines Erachtens ebenfalls die These der Unbestimmtheit: Auch diese Sätze bleiben durch Daten und Fakten nicht bloß unterbestimmt; bei ethischen und rechtlichen Normsätzen gibt es ebensowenig wie bei Übersetzungen Fakten, von denen sie „bestimmt" werden könnten. (Ich kann hier auf diese These, die meinen weiteren Ausführungen zugrunde liegt, nicht näher eingehen und verweise daher nur auf einige frühere Arbeiten, in denen ich sie vertreten und zu begründen versucht habe: Morscher (1973), (1974, 332–338), (1978), (1981), (1982).) Es wäre aber verfehlt, das Kind mit dem Bad auszuschütten und bei den normativen Sätzen in Recht und Ethik wegen ihrer Unbestimmtheit ganz auf Rationalität zu verzichten. Vielmehr müssen wir aus verschiedenen Gründen gerade für die Normen in diesen Bereichen den höchsten Rationalitätsstandard anstreben,

B. Brogaard, B. Smith (Hg.), *Rationality and Irrationality / Rationalität und Irrationalität*, S. 199–223.

der mit ihrer Unbestimmtheit vereinbar ist.

Diese abstrakten Überlegungen werde ich anhand eines konkreten Beispiels erläutern. Die Darstellung dieses Beispiels bildet den Hauptteil des vorliegenden Beitrages. Das Beispiel soll für einen konkreten Ausschnitt aus dem Rechtsleben zeigen, wie Rationalität in der Praxis des Rechts zum Tragen kommt. Daraus können gewisse Schlüsse bezüglich der Rationalität für das Recht im allgemeinen und darüber hinaus auch für die Ethik gezogen werden. Diese Schlußfolgerungen werden in einem abschließenden Abschnitt dieses Beitrages in aller Kürze kommentiert. Bevor ich jedoch das Beispiel – gewissermaßen als Fallstudie – präsentiere, muß ich eine terminologische Klarstellung bezüglich meiner Verwendung der Termini ‚Recht' und ‚Ethik' vorausschicken.

Recht und Moral sind gemäß der Terminologie, die ich hier verwende, Bestandteile unseres sozialen Lebens und der sozialen Wirklichkeit. Von dieser sozialen Wirklichkeit müssen wir theoretische Auseinandersetzungen und Reflexionen unterscheiden, die sich mit ihr beschäftigen und sie dadurch zu ihrem Untersuchungsgegenstand machen. In weitgehender Übereinstimmung mit Hans Kelsen verwende ich also die Termini ‚Recht' und ‚Moral' für diejenige soziale Wirklichkeit, von welcher die Rechtslehre und die Morallehre handeln. Den Terminus ‚Ethik' verwende ich im Sinne von ‚Morallehre': Ebenso wie das Recht der Untersuchungsgegenstand der Rechtslehre, so ist die Moral der Gegenstand der Ethik. Je nachdem, ob Recht und Moral rein deskriptiv untersucht werden oder ob man dazu auch normativ Stellung nimmt, kann man zwischen einer deskriptiven und einer normativen Rechts- und Morallehre unterscheiden. Eine wissenschaftliche Rechts- und Morallehre muß sich aber – und auch darin stimme ich mit Hans Kelsen überein – auf eine bloße Beschreibung und Erklärung der rechtlichen und moralischen Wirklichkeit beschränken und rein deskriptiv vorgehen, sie muß sich also jeder normativen Beurteilung bzw. Bewertung dieser sozialen Realität enthalten. Wissenschaftliche Rechtslehre bzw. Rechtswissenschaft und wissenschaftliche Morallehre bzw. wissenschaftliche Ethik sind somit rein deskriptive Disziplinen. Neben der rein deskriptiven Rechts- und Moralwissenschaft hat aber auch die normative Rechts- und Morallehre eine wichtige Aufgabe zu erfüllen. Eine solche normative Rechts- und Morallehre ist zwar nie rein wissenschaftlich möglich, entzieht sich deswegen jedoch noch lange nicht allen Standards an Rationalität. Mir geht es hier um die Frage, ob und in welchem Grad Rationalität in der Praxis des Rechtslebens – also im Recht – sowie in der normativen Rechtslehre und in der normativen Ethik erreichbar und wünschenswert ist.

2. Fallstudie (aus dem österreichischen Strafrecht)

a) *Der Sachverhalt*: Nehmen wir an, daß ein erwachsener Mann – nennen wir ihn ‚*a*' – einem anderen erwachsenen Mann – der Einfachheit halber ‚*b*' genannt – eine Überdosis eines Medikaments verabreicht und ihn dadurch tötet. Die Erhebungen der Exekutive und der gerichtsmedizinische Befund decken sich mit dem Geständnis von *a*. Der zentrale Satz im Protokoll, das der Staatsanwaltschaft übermittelt wird, lautet, daß *a* den *b* zum Zeitpunkt $t°$ getötet hat, in abgekürzter Schreibweise: $Tabt°$.

Für diesen Fall ist im österreichischen Strafgesetzbuch eine ganz einfache Rechtsnorm vorgesehen, die folgendermaßen lautet:

> § 75. Wer einen anderen tötet, ist mit Freiheitsstrafe von zehn bis zwanzig Jahren oder mit lebenslanger Freiheitsstrafe zu bestrafen.

Wer mit der Sprache des Rechts vertraut ist, weiß, daß ‚wer' am Anfang dieses Satzes als ‚wer auch immer' und damit als Allquantor zu lesen ist, und daß das ‚ist zu bestrafen' im normativen Sinn von ‚soll bestraft werden' zu verstehen ist. In einer standardisierten Sprache wird man daher den Satz ohne Änderung seines Sinnes folgendermaßen wiedergeben können:

> Für jeden Menschen *x* und für jeden Menschen *y* und für jeden Zeitpunkt *t* gilt: Wenn *x* den *y* zu *t* tötet, dann ist es gesollt, daß *x* für das, was *x* zu *t* getan hat, mit einer Freiheitsstrafe von zehn bis zwanzig Jahren oder mit lebenslanger Freiheitsstrafe bestraft wird.

Mit Hilfe von z.T. gängigen und z.T. naheliegenden Abkürzungen können wir (wenn wir wollen) diesen Satz besonders übersichtlich in einer symbolischen Logik-Sprache wiedergeben. Wir benützen dabei folgende Symbole zur Abkürzung:

> den Allquantor ‚$\forall x$' für: ‚für jeden Menschen *x* gilt' (und analog ‚$\forall y$' für: ‚für jeden Menschen *y* gilt', und ‚$\forall t$' für: ‚für jeden Zeitpunkt *t* gilt')
> ‚$\rightarrow$' für: ‚wenn – dann'
> ‚$\vee$' für: ‚oder'
> ‚O' für: ‚es ist gesollt (bzw. geboten), daß'

Außerdem führen wir neben dem dreistelligen Prädikat ‚$Txyt$' für ‚*x* tötet *y* zum Zeitpunkt *t*' (bzw. ‚*x* hat *y* zum Zeitpunkt *t* getötet') noch zusätzlich folgende zweistellige Prädikate ein:

> ‚B_1xt' für: ‚*x* wird für das, was *x* zu *t* getan hat, mit einer Freiheitsstrafe von zehn bis zwanzig Jahren bestraft'
> ‚B_2xt' für: ‚*x* wird für das, was *x* zu *t* getan hat, mit lebenslanger Freiheitsstrafe bestraft'

§ 75 des österreichischen Strafgesetzbuches erhält mit Hilfe dieser Abkürzungen folgende Form:

$$\forall x \forall y \forall t (Txyt \rightarrow O(B_1xt \vee B_2xt))$$

b) Der *Staatsanwalt* könnte nun den Fall vorläufig folgendermaßen aufbereiten (wobei ‚*a*' eine Abkürzung für den Namen des Täters, ‚*b*' eine Abkürzung für den Namen des Opfers und ‚$t°$' die Bezeichnung für den Zeitpunkt der Tat sei):

1.	$Tabt°$	Tatsachenbehauptung
2.	$\forall x \forall y \forall t (Txyt \rightarrow O(B_1xt \vee B_2xt))$	§ 75 StGB
3.	$Tabt° \rightarrow O(B_1at° \vee B_2at°)$	aus (2) durch 3malige Anwendung von US
4.	$O(B_1at° \vee B_2at°)$	aus (1) und (3) durch MP

US ist die Regel der Universellen Spezifikation, der zufolge man von einem Satz der Form ‚$\forall x(\ldots x\ldots)$' zum Satz ‚$\ldots a\ldots$' und analog von ‚$\forall y(\ldots y\ldots)$' zu ‚$\ldots b\ldots$' und von ‚$\forall t(\ldots t\ldots)$' zu ‚$\ldots t^{\circ}\ldots$' übergehen darf. MP ist der bekannte Modus Ponens, der den Übergang von zwei Sätzen A und $A \rightarrow B$ zu B gestattet.

Spätestens hier wird sichtbar, daß ich absichtlich eine möglichst einfache Art der formalen Darstellung wähle und auf eine differenziertere Formalisierung (insbesondere auch im Hinblick auf bedingte Normsätze, etwa mit Hilfe von dyadischen Normoperatoren) verzichte, weil diese elementare Formalisierung für die Fragestellung, um die es mir geht, genügt.

c) Nun tritt der *Verteidiger* von a auf den Plan. Er wird versuchen, a zu „entschuldigen". Dies ist auf verschiedene Art und Weise möglich. So kann er z.B. versuchen, das Faktum selbst, daß a den b getötet hat, schlicht zu bestreiten und somit nachzuweisen, daß *nicht* zutrifft: $Tabt^{\circ}$. In unserem Fall wird er damit nicht weit kommen, da wir ja davon ausgegangen sind, daß ein Geständnis des Täters vorliegt, mit dem die Erhebungen und der gerichtsmedizinische Befund übereinstimmen. Es bleiben dem Anwalt aber noch eine Reihe anderer Möglichkeiten offen; auf zwei besonders wichtige und typische Strategien, den Beschuldigten zu entlasten, möchte ich näher eingehen.

Zunächst einmal könnte der Verteidiger in Zweifel ziehen, daß die Handlung von a – von der wir hier außer Streit stellen, daß sie geschehen ist – überhaupt rechtswidrig und schuldhaft war (Verteidigungsstrategie A). Unabhängig davon könnte der Verteidiger aber auch in Frage stellen, ob die Art, wie a den b getötet hat, überhaupt nach § 75 StGB zu qualifizieren und zu behandeln ist (Verteidigungsstrategie B).

Verteidigungsstrategie A: Wenn jemand, der einen anderen tötet, in Notwehr gehandelt hat, hätte der Verteidiger die Möglichkeit, die Anwendung von § 75 StGB auf diesen Fall zu bekämpfen, weil nach § 3 StGB jemand, der in Notwehr handelt, nicht rechtswidrig handelt. Da die beispielhafte Beschreibung unseres Falles Notwehr praktisch ausschließt, wollen wir hier von dieser Möglichkeit absehen. Ein anderer Paragraph des StGB könnte jedoch für die Zwecke der Verteidigung durchaus brauchbar sein, nämlich § 4 StGB, der besagt:

> § 4. Strafbar ist nur, wer schuldhaft handelt.

Der in § 4 verwendete Begriff der Strafbarkeit ist offenbar ein normativer Begriff: Daß eine Handlung strafbar ist, heißt demnach nicht, daß sie bestraft werden *kann*, sondern daß sie bestraft werden *darf*, daß es also *erlaubt* ist, sie zu bestrafen. Daß jemand nur strafbar ist, wenn er schuldhaft handelt, besagt somit:

> Für jeden Menschen x und für jeden Zeitpunkt t gilt: Wenn x zu t nicht schuldhaft handelt bzw. gehandelt hat, dann ist es nicht erlaubt, x für das, was x zu t getan hat, zu bestrafen.

Zur vereinfachten Wiedergabe dieses Satzes führen wir einige weitere Abkürzungen ein, und zwar:

‚$\neg$' für: ‚nicht' (bzw. ‚es ist nicht der Fall, daß')
‚Sxt' für: ‚x handelt zu t schuldhaft' (bzw. ‚x hat zu t schuldhaft gehandelt')
‚Bxt' für: ‚x wird für das, was x zu t getan hat, bestraft'
‚P' für: ‚es ist erlaubt, daß'

§ 4 StGB lautet dann in abgekürzter Schreibweise:

$$\forall x\forall t(\neg Sxt \rightarrow \neg P(Bxt))$$

Um diesen Paragraphen für unseren Fall nutzen zu können, müßte der Anwalt noch einen weiteren Paragraphen heranziehen – z.B. § 11 StGB, der besagt, daß jemand, der zum Zeitpunkt einer Tat nicht zurechnungsfähig war, nicht schuldhaft gehandelt hat, d.h.:

> Für jeden Menschen x und für jeden Zeitpunkt t gilt: Wenn x zu t nicht zurechnungsfähig ist, dann handelt x zu t nicht schuldhaft.

Zur formalen Darstellung dieses Satzes benötigen wir noch zusätzlich die Abkürzung ‚Zxt' für ‚x ist zu t zurechnungsfähig'. Wir erhalten dann für § 11 StGB folgende Formulierung:

$$\forall x\forall t(\neg Zxt \rightarrow \neg Sxt)$$

Nun kann der Verteidiger unter der Voraussetzung, daß a zu t° nicht zurechnungsfähig war ($\neg Zat^{\circ}$), durch folgende Argumentation begründen, daß – entgegen der Schlußfolgerung des Staatsanwaltes – nicht zutrifft, daß a zu einer Freiheitsstrafe von zehn bis zwanzig Jahren oder zu einer lebenslänglichen Freiheitsstrafe verurteilt werden soll; ich verwende dabei auch noch das Symbol ‚$\Box$' für die Modalphrase ‚es ist notwendig, daß':

1.	$\neg Zat^{\circ}$	Tatsachenbehauptung
2.	$\forall x\forall t(\neg Sxt \rightarrow \neg P(Bxt))$	§ 4 StGB
3.	$\forall x\forall t(\neg Zxt \rightarrow \neg Sxt)$	§ 11 StGB
4.	$\neg Zat^{\circ} \rightarrow \neg Sat^{\circ}$	aus (3) durch 2malige Anwendung von US
5.	$\neg Sat^{\circ}$	aus (1) und (4) durch MP
6.	$\neg Sat^{\circ} \rightarrow \neg P(Bat^{\circ})$	aus (2) durch 2malige Anwendung von US
7.	$\neg P(Bat^{\circ})$	aus (5) und (6) durch MP
8.	$O(\neg Bat^{\circ})$	aus (7) durch R1
9.	$\Box(\neg Bat^{\circ} \rightarrow \neg(B_1at^{\circ} \vee B_2at^{\circ}))$	*per definitionem* von ‚B', ‚B_1' und ‚B_2'
10.	$O\neg(B_1at^{\circ} \vee B_2at^{\circ})$	aus (8) und (9) durch R2
11.	$\neg O(B_1at^{\circ} \vee B_2at^{\circ})$	aus (10) durch R3

Die drei Regeln R1, R2 und R3 sind Regeln der deontischen Logik bzw. Normenlogik, die in allen deontischen Standardsystemen gelten:

R1 erlaubt den Übergang von ‚$\neg P(p)$' (‚es ist nicht erlaubt, daß p') zu ‚$O(\neg p)$' (‚es ist geboten, daß nicht p') – es handelt sich dabei nur um zwei verschiedene Formulierungen des Verbotes von p.

Die Regel R2 gestattet den Übergang von einem Sollsatz ‚$O(p)$' und der deskriptiven Prämisse ‚$\Box(p \rightarrow q)$' zu ‚$O(q)$'. Diese Regel ist in den Standardsystemen der deontischen Logik, in deren Sprache auch alethische Modaloperatoren wie ‚$\Box$' enthalten sind, gültig. (Man kann die Regel R2 in diesen Standardsystemen im allgemeinen auf einfachere Regeln bzw. Gesetze – nämlich auf ‚$\Box(p) \rightarrow O(p)$' und ‚$O(p \rightarrow q) \rightarrow (Op \rightarrow Oq)$' – zurückführen.) Die Regel R2 ist mit einem Prinzip verwandt, das am Beginn der Geschichte der Imperativ- und Normenlogik in den 30er Jahren des 20.Jahrhunderts bei der Suche nach einer Semantik für Imperativ- und Normsätze eine wichtige Rolle spielte. Es besagt: Wann immer die Erfüllung einer Norm oder eines Imperativs N_2 aus der Erfüllung einer anderen Norm bzw. eines anderen Imperativs N_1 logisch folgt, dann folgt auch N_2 selbst aus N_1. Man nannte die auf diesem Prinzip basierende Logik ‚Erfüllungslogik' bzw. ‚Logic of Satisfaction' (vgl. Ross (1941); ich werde im Appendix darauf näher eingehen). Der Grundsatz der Erfüllungslogik besagt nicht genau dasselbe wie unsere Regel R2; wegen ihrer offenkundigen Verwandtschaft mit diesem Grundsatz werde ich jedoch die Regel R2 im folgenden gelegentlich auch als Regel der Erfüllungslogik bzw. der „Logic of Satisfaction" apostrophieren.

R3 schließlich beruht auf dem bekannten normenlogischen Widerspruchsprinzip, das ausschließt, daß etwas zugleich verboten ($O(\neg p)$) und geboten ($O(p)$) sein kann. R3 gestattet somit den Übergang von ‚$O(\neg p)$' zu ‚$\neg O(p)$'. Aus der Regel R3 folgt, daß das, was geboten ist, immer auch erlaubt sein muß; das ergibt sich aufgrund der Definition der Erlaubnis ($P(p) :\leftrightarrow \neg O(\neg p)$) durch Kontraposition der Regel R3 oder auch durch Substitution von ‚$\neg p$' für ‚p'.

Wird dem Verteidiger vom Gericht die Tatsachenbehauptung, daß a zu $t°$ unzurechnungsfähig war ($\neg Zat°$), nicht abgenommen, bleibt ihm immer noch die Möglichkeit, zumindest das Strafausmaß, das a für seine Handlung aufgrund von § 75 StGB droht, herabzumindern. Dies führt uns zu Strategie B.

Verteidigungsstrategie B: Betrachtet man § 75 StGB genauer, stellt man als Laie überrascht fest, daß die Überschrift „Mord" lautet, obwohl im Text von § 75 nur von ‚töten' die Rede ist. Der Jurist weiß natürlich von allem Anfang an, auf welche allgemeinen Einschränkungen und Voraussetzungen es dabei ankommt, die nicht bei den einzelnen Paragraphen jeweils wiederholt werden, wie z.B.

> § 7. (1) Wenn das Gesetz nichts anderes bestimmt, ist nur vorsätzliches Handeln strafbar.

Ebenso weiß der Jurist, daß der § 75 (Mord) gar nicht zur Anwendung kommt, wenn es sich um eine besondere Art des Tötens handelt bzw. wenn das Töten unter besonderen Umständen erfolgt wie z.B. im Falle von Totschlag (§ 76), Töten auf Verlangen (§ 77) oder Fahrlässiger Tötung (§ 80). Nehmen wir als Beispiel die Tötung auf Verlangen:

> § 77. Wer einen anderen auf dessen ernstliches und eindringliches Verlangen tötet, ist mit Freiheitsstrafe von sechs Monaten bis zu fünf Jahren zu bestrafen.

Wir können diesen Satz zunächst wiedergeben durch:

> Für jeden Menschen x und für jeden Menschen y und für jeden Zeitpunkt t gilt: Wenn y von x zu t ernstlich und eindringlich verlangt, getötet zu werden, und x den y zu t tötet, dann ist es geboten, daß x mit einer Freiheitsstrafe von sechs Monaten bis zu fünf Jahren bestraft wird.

Nun führen wir wieder ein paar Abkürzungen ein, nämlich:

> ‚$Eyxt$' für: ‚y verlangt von x zu t ernstlich und eindringlich, getötet zu werden'
> ‚B_3xt' für: ‚x wird für das, was x zu t getan hat, mit einer Freiheitsstrafe von sechs Monaten bis zu fünf Jahren bestraft'
> ‚$\wedge$' für: ‚und'

Mit Hilfe dieser und der bereits zuvor eingeführten Abkürzungen können wir § 77 StGB folgendermaßen wiedergeben:

$$\forall x \forall y \forall t((Eyxt \wedge Txyt) \rightarrow O(B_3xt))$$

Sobald ein Tatbestand unter eine speziellere Rechtsnorm fällt, wird die allgemeinere Rechtsnorm, unter die er ebenfalls fällt, „verdräng", und sie kommt dann – wegen des Grundsatzes „ne bis in idem" – nicht mehr zur Anwendung. (Auf die Relevanz dieses Grundsatzes im vorliegenden Kontext hat mich dankenswerterweise Andrew U. Frank aufmerksam gemacht.) Der Anwalt könnte daher eine neue Verteidigungslinie aufbauen, wenn er erkennt, daß a nicht straffrei bleibt; um a wenigstens die höhere Strafe nach § 75 StGB zu ersparen, könnte er folgendermaßen argumentieren:

1.	$Ebat° \wedge Tabt°$	Tatsachenbehauptung
2.	$\forall x \forall y \forall t((Eyxt \wedge Txyt) \rightarrow O(B_3xt))$	§ 77 StGB
3.	$(Ebat° \wedge Tabt°) \rightarrow O(B_3at°)$	aus (2) durch 3malige Anwendung von US
4.	$O(B_3at°)$	aus (1) und (3) durch MP
5.	$\forall x \forall t(O(B_3xt) \rightarrow O\neg(B_1xt \vee B_2xt))$	gemäß dem Grundsatz „ne bis in idem"
6.	$O(B_3at°) \rightarrow O\neg(B_1at° \vee B_2at°)$	aus (5) durch 2malige Anwendung von US
7.	$O\neg(B_1at° \vee B_2at°)$	aus (4) und (6) durch MP
8.	$\neg O(B_1at° \vee B_2at°)$	aus (7) durch R3

Das hier geschilderte Vorgehen entspricht der Art und Weise, wie die Rechtsnormen im StGB aufgeführt sind. Dieses Vorgehen stimmt überraschend weitgehend mit den Schlußverfahren überein, die uns vom „non-monotonen Schließen" und von der „Default Logic" bzw. dem „Defeasible Reasoning" her bekannt sind. Mit der von mir gewählten Art der Darstellung des Fallbeispiels wollte ich diese Ähnlichkeiten und Übereinstimmungen herausarbeiten. Dabei betrachten wir § 75 StGB zunächst einmal als *prima facie* anzuwendende Regel, d.h. als Regel, die anzuwenden ist, solange nichts Gegenteiliges bekannt ist und kein Einspruch erfolgt. Werden nähere faktische Voraussetzungen (wie z.B. Unzurechnungsfähigkeit von a) oder auch bestimmte Qualifikationen des Tatbestandes (wie z.B. allgemein verständliche Gemütsbewegung des Täters oder ernstliches und eindringliches Verlangen des Opfers, getötet zu werden) bekannt, kann die allgemeine

Rechtsnorm des § 75 StGB „defeatet" werden – d.h. sie wird von anderen Rechtsnormen „verdrängt" und kommt dadurch im vorliegenden Fall nicht mehr zur Anwendung.

Natürlich hätte man für das hier behandelte Beispiel auch eine Darstellung wählen können, bei der die – für den Juristen selbstverständlichen – Vorbedingungen vollständig aufgezählt werden. Wir fügen zu diesem Zweck unserem Vokabular noch folgende Abkürzungen hinzu:

‚*Nxt*' für: ‚*x* handelt zu *t* in Notwehr' (§ 3)
‚*Sxt*' für: ‚*x* handelt zu *t* schuldhaft' (§ 4; diese Abkürzung wurde schon früher verwendet)
‚*Vxt*' für: ‚*x* handelt zu *t* vorsätzlich' (§ 7 (1))
‚*Wxt*' für: ‚*x* hat zu *t* das zwanzigste Lebensjahr vollendet' (§ 36)
‚*Gxt*' für: ‚*x* handelt zu *t* in einer allgemein begreiflichen heftigen Gemütsbewegung' (§ 76)

Dann könnten wir § 75 StGB etwa folgendermaßen vervollständigen, indem wir die stillschweigend hinzugedachten (und für den Juristen „selbstverständlichen") Bedingungen ergänzen und explizit anführen:

$$\forall x \forall y \forall t((Txyt \wedge \neg Nxt \wedge Sxt \wedge Vxt \wedge Wxt \wedge \neg Gxt \wedge \neg Eyxt) \rightarrow O(B_1x \vee B_2x))$$

In diesem Fall betrachtet man das ursprünglich vorgeschlagene Argument als bloßes Enthymem: Die bisherige normative Prämisse (§ 75 StGB) muß dabei durch Hinzufügung neuer Bedingungen vervollständigt werden. Eine derart vervollständigte Rechtsnorm kann und muß nicht mehr „verdrängt" werden. Damit das Argument logisch korrekt bleibt, sind jedoch zusätzliche Prämissen erforderlich, welche den ergänzten bzw. neuen Antezedensbedingungen der Rechtsnorm entsprechen. Die Verteidigungsstrategie wird in diesem Fall darauf hinauslaufen, eine der neuen Prämissen als falsch zu entlarven und dadurch dem Argument die Beweiskraft zu nehmen. Die Gefahr bei dieser Art der „Aufbereitung" eines Falles besteht darin, daß die angestrebte Vollständigkeit meist nicht erreicht wird und daß Vorbedingungen, auf die der Gesetzgeber vergessen hat oder die zur Zeit der Gesetzgebung gar nicht in Betracht kamen, nicht einfach *ad hoc* hinzugefügt werden können.

Abgesehen von diesen praktischen Unterschieden, handelt es sich jedoch bloß um zwei verschiedene Darstellungsweisen, die theoretisch gleichwertig sind.

d) Nach all diesen Vorüberlegungen ist der *Richter* an der Reihe. Nehmen wir einmal an, er lasse sich von den Argumenten des Verteidigers nicht beeindrucken und schließe sich dem ursprünglichen Argument des Staatsanwaltes an. Er schließe also wie dieser:

1. $Tabt^\circ$
2. $\forall x \forall y \forall t(Txyt \rightarrow O(B_1xt \vee B_2xt))$
3. $Tabt^\circ \rightarrow O(B_1at^\circ \vee B_2at^\circ)$
4. $O(B_1at^\circ \vee B_2at^\circ)$

Nehmen wir nun ferner an, der Richter verurteile *a* zu lebenslanger Freiheitsstrafe, sein Urteil laute also:

$O(B_2at^\circ)$

Wie kann der Richter dieses Urteil durch ‚$O(B_1at^\circ \vee B_2at^\circ)$' begründen bzw. aus ‚$O(B_1at^\circ \vee B_2at^\circ)$' ableiten? Nach Kelsen hängt die Begründung davon ab, daß der Inhalt der singulären Norm, die der Richter durch sein Urteil setzt, dem Inhalt der allgemeinen Norm „entspricht" (Kelsen (1968), (1979), 208 ff.). Was aber heißt das?

Lösung A: Eine erste Antwort auf diese Frage lautet, daß wir ‚$O(B_2at^\circ)$' „direkt" aus der Konklusion ‚$O(B_1at^\circ \vee B_2at^\circ)$', die Richter und Staatsanwalt aus Tatsachenbehauptungen und Rechtsnormen logisch korrekt erschlossen haben, ableiten können, und zwar folgendermaßen:

⋮

4.	$O(B_1at^\circ \vee B_2at^\circ)$	Konklusion des vorausgehenden Arguments
5.	$\Box(B_2at^\circ \rightarrow (B_1at^\circ \vee B_2at^\circ))$	elementares modallogisches Gesetz
6.	$O(B_2at^\circ)$	aus (4) und (5) – aber wie?

Der Übergang von (4) und (5) wird dabei durch eine spezielle Regel gerechtfertigt, welche in gewissem Sinn die Umkehrung unserer früheren Regel R2 ist und besagt: Von ‚$O(p)$' und ‚$\Box(q \rightarrow p)$' darf man übergehen zu ‚$O(q)$'. Aus der Sicht der deontischen Standardlogik handelt es sich beim Übergang von (4) und (5) zu (6) in der obigen Ableitung schlicht um einen Fehlschluß – vergleichbar dem „Modus Morens", wonach man – in „Verdrehung" des Modus Ponens – von einem Wenn-Dann-Satz und seinem Konsequens auf sein Antezedens schließt. Eine Schlußregel, die den Übergang von (4) und (5) zu (6) und damit von ‚$O(p)$' und ‚$\Box(q \rightarrow p)$' auf ‚$O(q)$' legitimiert, macht aus einer normativen Begründungsnot eine logische Tugend. Selbst große Philosophen der Vergangenheit sind diesem Fehlschluß erlegen, und auch noch im 20.Jahrhundert wurde er immer wieder mehr oder weniger explizit als korrekter Schluß einer „praktischen Syllogistik" oder eines „praktischen Schließens" verteidigt. Anthony Kenny gründete auf eine solche Schlußregel sogar eine eigene Imperativ-Logik, der er – im Gegensatz zur „Logic of Satisfaction" – den Namen „Logic of Satisfactoriness" gab (vgl. Kenny (1966), (1976), 80 ff.; auch darauf gehe ich im Appendix näher ein).

Die „Schlußregel", durch welche der Fehlschluß von (4) und (5) auf (6) gerechtfertigt würde, wurde im vorliegenden Kontext genau zu dem Zweck *ad hoc* eingeführt, um das Problem des Richters zu lösen, wie er sein Urteil rechtfertigen kann. (Der ganzen „Logic of Satisfactoriness" liegt wohl ein ähnlich edles Motiv zugrunde.) Aber diese „Schlußregel" taugt nicht einmal zur Lösung dieses Problems selbst. Das ergibt sich daraus, daß mit genau demselben Recht wie ‚$O(B_2at^\circ)$' durch diese Schlußregel auch ‚$O(B_1at^\circ)$' abgeleitet und damit begründet werden könnte. Das Urteil ‚$O(B_2at^\circ)$' wird daher durch diese Ableitung überhaupt nicht begründet, sondern erweist sich als völlig willkürlich, wenn wir es nur mit dieser Regel rechtfertigen könnten. Kurz: Die „Logic of Satisfactoriness" ist sowohl „from a logical point of view" als auch „from a legal point of view" nicht „satis-

factory". (Sollte jemand diesen Einwand mit dem Hinweis zu bagatellisieren versuchen, daß es bei einem Menschen meines Alters ja ohnedies keinen allzu großen Unterschied mehr ausmacht, ob er mit zehn- bis zwanzigjährigem oder aber mit lebenslänglichem Freiheitsentzug bestraft wird, so kann das Beispiel jederzeit so zugespitzt werden, daß es niemandem mehr gleichgültig ist: In manchen – sogar „zivilisierten" – Ländern wird ja auch heute noch die Todesstrafe verhängt und sogar vollstreckt; es dürfte aber niemandem gleichgültig sein, ob er für den Rest seines Lebens im Gefängnis sitzt oder für den Rest seines Lebens tot ist ...)

Lösung B: Wir müssen infolgedessen einen anderen Lösungsansatz suchen. Dabei müssen wir Gründe angeben, welche die alternative Bestrafung und damit ‚$B_1at°$' ausschließen. Das Gesetz gibt allgemeine Grundsätze dafür an, wie die jeweilige Strafe zu bemessen ist (§§ 32 ff. StGB). Diese Grundsätze sind heranzuziehen, um durch Plausibilitätsargumente, Wahrscheinlichkeitsschlüsse, Abwägungen etc. eine Prämisse folgender Art zu begründen:

$O(\neg B_1at°)$

Diese Voraussetzung ist selbst die Konklusion einer sehr komplizierten Argumentation und immer nur mit gewisser Wahrscheinlichkeit und Plausibilität zu rechtfertigen. Steht dem Richter aber eine solche Prämisse zur Verfügung, kann er sein Urteil damit nun streng logisch (nämlich gemäß den Gesetzen der „Standardlogik" und der „Logic of Satisfaction") folgendermaßen begründen:

⋮		
4.	$O(B_1at° \vee B_2at°)$	Konklusion des früheren Arguments
5.	$O(\neg B_1at°)$	auf Gesetzen beruhende und mit Tatsachenbehauptungen abzustützende Voraussetzung
6.	$O(B_1at° \vee B_2at°) \wedge O(\neg B_1at°)$	aus (4) und (5) durch Konjunktion
7.	$O[(B_1at° \vee B_2at°) \wedge \neg B_1at°]$	aus (6) durch R5
8.	$\Box[((B_1at° \vee B_2at°) \wedge \neg B_1at°) \rightarrow B_2at°]$	elementares modallogisches Gesetz
9.	$O(B_2at°)$	aus (7) und (8) durch R2

Die dabei zusätzlich verwendete Regel R5 ist in allen Standardsystemen der deontischen Logik gültig: Sie gestattet den Übergang von ‚$O(p) \wedge O(q)$' zu ‚$O(p \wedge q)$'.

Die wichtigste Regel, auf der diese Ableitung beruht, ist aber R2, die – zum Unterschied von ihrer „Verdrehung" in der „Logic of Satisfactoriness" – logisch „einwandfrei" ist. Die Hauptarbeit der Urteilsbegründung liegt – diesem Modell zufolge – in der Begründung der Voraussetzung (5). Diese Arbeit nimmt uns die Logik selbst nicht ab – wie die Logik ja weder imstande ist noch beansprucht, allein Probleme lösen zu können; durch die Klärung von Problemen leistet die Logik allerdings einen wesentlichen *Beitrag* zu ihrer Lösung.

3. Kommentar

Bei der Fallstudie aus dem österreichischen Strafrecht, die ich im vorausgehenden Abschnitt vorgestellt habe, ging es mir nicht darum, ein strafrechtliches Verfahren zu beschreiben. Ich wollte damit vielmehr beispielhaft aufzeigen, wie sich Rationalität im rechtlich-normativen Diskurs darbietet und daß man auch in einer normativen Auseinandersetzung nicht auf Rationalität verzichten kann und darf. Die Übertragbarkeit der charakteristischen Merkmale dieser Fallstudie auf den ethischen Diskurs liegt auf der Hand; der einzige wesentliche Unterschied liegt darin, daß wir in der ethischen Diskussion die normativen Prinzipien nicht einem Kodex von (positiv gesetzten) Normen entnehmen können.

Zwei entgegengesetzte Vorurteile überschatten die gängigen Auffassungen von Rationalität. Einerseits wird häufig unterstellt, Rationalität sei nur im Rahmen der deduktiven Logik möglich und daher auf Regeln der deduktiven Logik angewiesen (a). Anderseits entsteht oft der Eindruck, Rationalität sei zwar an Regeln gebunden, aber jede beliebige Art von Regeln gewährleiste bereits Rationalität (b). Diese beiden Vorurteile hatte ich bereits bei der Darbietung der Fallstudie im Visier; im folgenden werde ich sie nochmals separat kommentieren. Außerdem werde ich mich aber auch noch mit einem weiteren Vorurteil auseinandersetzen; es geht auf Aristoteles zurück und besagt, daß nur die Mittel für bestimmte Zwecke, niemals aber diese Zwecke selbst rational diskutiert oder gar begründet werden können (c). Schließlich werde ich auch noch der Frage nachgehen, welche Stufe an Rationalität in Recht und Ethik überhaupt wünschenswert ist – und warum (d).

a) Die im zweiten Abschnitt anhand der Fallstudie aus dem österreichischen Strafrecht rekonstruierte Argumentationsweise erinnert frappant an das Vorgehen im Rahmen des „Defeasible Reasoning" oder einer „non-monotonen Logik". Die Rationalität der rechtlichen Argumentation leidet darunter nicht. Rationalität setzt also nicht deduktives Denken oder gar eine bestimmte Form der deduktiven Logik voraus, sondern ist mit einer nicht-deduktiven Logik (wie z.B. einem non-monotonen Schließen oder einem „Defeasible Reasoning") durchaus vereinbar. Ja, diese Art des Schließens scheint sich für den Bereich des Rechts und für die Rekonstruktion rechtlicher Argumentationen besonders gut zu eignen (und Analoges gilt für den ethischen Diskurs). Das hängt mit einer charakteristischen Eigenschaft des Rechts selbst und der Gesetzgebung zusammen: mit ihrer Offenheit („open texture") bzw. Unbestimmtheit („indeterminacy"); diese ist ihrerseits durch zwei Grundzüge des Menschen bedingt: sein eingeschränktes Tatsachenwissen und die relative Unbestimmtheit seiner Ziele (Hart (1961), 124 ff.).

Angesichts der weiten Verbreitung und geradezu „natürlichen" Anwendung der Methode des „Defeasible Reasoning" im Bereich des Rechts braucht es einen nicht wunderzunehmen, daß sich einer der frühesten Belege für die Verwendung des Terminus ‚defeasible' in der philosophischen Literatur ausgerechnet im Werk des prominenten englischen Rechtsphilosophen H.L.A. Hart findet (vgl. Donald Nute in seiner „Preface" zu Nute (1997), VII). Manche Autoren sprachen sogar von einer besonderen „defeasability of practical reasoning" (Geach (1966), 78) und behaupteten, daß das praktische Schließen insgesamt in einer Weise „defeasible" sei, in der dies auf das theoretische Schließen nicht

zutreffe (Geach (1966), 77). Wenn dies aus heutiger Sicht auch übertrieben erscheint, so sollte man doch angesichts der zahlreichen lebensnahen Anwendungsfälle aus dem Bereich des praktischen Schließens, insbesondere auch aus dem Bereich des Rechts und der Ethik, allmählich daran denken, die immer wieder strapazierten und manchmal recht verkrampft wirkenden Standardbeispiele des non-monotonen Schließens (wie: Tweety ist ein Vogel und müßte daher normalerweise fliegen können; nun ist aber Tweety ein Pinguin und kann daher nicht fliegen) durch realistischere Beispiele zu ersetzen wie z.B.: *a* hat *b* getötet und sollte daher normalerweise nach § 75 StGB mit zehn bis zwanzig Jahren Freiheitsstrafe oder lebenslanger Freiheitsstrafe bestraft werden; nun hat aber *b* von *a* ernsthaft und eindringlich verlangt, getötet zu werden, und daher soll *a* nicht mit zehn- bis zwanzigjähriger oder gar lebenslänglicher Freiheitsstrafe, sondern bloß mit einer Freiheitsstrafe von sechs Monaten bis zu fünf Jahren bestraft werden.

b) Wenn Rationalität schon nicht an die deduktive Logik gebunden ist, ist sie dann überhaupt noch an eine Logik gebunden? Manche Philosophen vertreten (im Anschluß an Hume) auch heute noch die Auffassung, daß Logik nur deduktiv oder defektiv sein kann, daß also eine Logik, die nicht deduktiv ist, automatisch defektiv sein muß und daher gar keine Logik ist. Wie weit oder wie eng man die Grenzen der Logik zieht, mag durch eine terminologische Festlegung entschieden werden. Wenn man aber Rationalität nicht auf deduktive Logik beschränkt, heißt das sicher nicht, daß man sie damit für völlig „vogelfrei" erklärt und daß sie bar jeder Logik sein kann. Eine Regel allein ist jedenfalls bei weitem noch kein Garant für Rationalität. Dazu ist es erforderlich, daß eine solche Regel bestimmte Bedingungen erfüllt, daß es sich dabei also zumindest in einem weiteren Sinn um eine „logische" Regel handelt.

Was aber müssen wir von den Regeln eines Regelsystems als Minimum verlangen, um von einer Logik und in weiterer Folge dann auch von Rationalität sprechen zu können? Die Regeln der deduktiven Logik zeichnen sich dadurch aus, daß sie, auf Aussagesätze angewandt, deren Wahrheit (und in gewissen Fällen sogar deren logische Wahrheit) „erhalten", d.h. von den Prämissen auf die Konklusion eines Arguments, das diesen Regeln entspricht, „übertragen". Eine Verallgemeinerung dieser Idee führt zum Gedanken, daß logische Regeln immer – wenn schon nicht die Wahrheit, so doch zumindest – eine analoge Eigenschaft erhalten, d.h. von den Prämissen auf die Konklusion übertragen müssen. Nicht jede beliebige Eigenschaft taugt dabei jedoch als Wahrheitsersatz: Eine Regel, die eine x-beliebige Eigenschaft von den Prämissen auf die Konklusion eines Arguments, das diesen Regeln entspricht, überträgt, macht noch keine Logik aus. (Man denke z.B. an die Eigenschaft, einen Anfangsbuchstaben aus der ersten Hälfte des Alphabets oder eine bestimmte minimale oder maximale Satzlänge – mehr als 10 Buchstaben oder weniger als 1000 Buchstaben – zu haben. Es ist leicht einzusehen, daß eine Regel, welche solche Eigenschaften überträgt, keine logische Regel ist.) Von welcher Art müssen aber diese Eigenschaften von Prämissen eines Arguments sein, damit sie logisch relevant sind und ihre Übertragung auf die Konklusion als Grundlage einer Logik dienen kann? Diese Frage stellt sich naturgemäß besonders dringlich bei Sätzen, die gar keine Wahrheitswerte haben können, die also gar nicht im üblichen Sinn des Wortes wahr oder falsch sein können, wie z.B. bei Imperativen und Normsätzen. Man hat daher gerade bei diesen Sätzen besonders intensiv nach einer Ersatzeigenschaft gesucht, die anstelle der Wahrheit treten und deren Übertragung von den Prämissen auf die Konklusion eines Arguments

eine Logik der Imperative und Normen begründen könnte. Als erste Lösung dieses Problems wurde – in den „Uranfängen" der Imperativlogik – vorgeschlagen, eine Schlußregel der Imperativlogik müsse gewährleisten, daß die Erfüllung der Imperative von den Prämissen auf die Konklusion übertragen wird; damit wurde die sogenannte „Erfüllungslogik" („Logic of Satisfaction") begründet. Die sogenannte „Logic of Satisfactoriness" setzte anstelle der „Satisfaction" die „Satisfactoriness" eines Imperativs bzw. einer Norm als denjenigen Wert, der durch eine Schlußregel erhalten werden muß und der damit der Imperativlogik zugrunde liegt. In unserer Fallstudie haben wir gesehen, daß die „Logic of Satisfactoriness" in rechtlichen Fragen nicht zu begründeten Entscheidungen führt, die sich von unbegründeten Entscheidungen klar unterscheiden, sondern daß sie die Urteilsfindung zu einer Geschmacksfrage degradiert; sie garantiert daher nicht Rationalität, sondern öffnet der Beliebigkeit Tür und Tor. Mit ihrer Grundschlußregel, welche den Übergang von ‚$O(p)$' und ‚$\Box(q \rightarrow p)$' zu ‚$O(q)$' legitimiert, erhebt sie die fragwürdige Maxime „Der Zweck heiligt die Mittel" zu einem logischen Gesetz. (Wie schon erwähnt, gehe ich im Appendix näher auf die „Logic of Satisfactoriness" ein.)

Daß Rationalität nur innerhalb der strengen Grenzen der deduktiven Logik möglich sei, ist ebenso verkehrt wie das Vorurteil, daß Rationalität ganz ohne Logik auskommen könne. Diese beiden Vorurteile, denen ich in a) und b) entgegengetreten bin, betreffen die Art und die Ausprägung von Rationalität, um die es im Bereich von Recht und Ethik geht. Es gibt aber auch ein Vorurteil, welches den Anwendungsbereich der Rationalität innerhalb von Recht und Ethik betrifft und auf das ich nunmehr kurz eingehen werde.

c) Einem geradezu „klassischen" Vorurteil zufolge kann man zwar über Mittel, nie aber über Ziele und Zwecke rational diskutieren: „Unsere Überlegung betrifft nicht das Ziel, sondern die Mittel, es zu erreichen" (Aristoteles, *Nikomachische Ethik* 1112b, 10). Bei der Darstellung der Fallstudie habe ich einfach die derzeit in Österreich geltenden Gesetze vorausgesetzt und aufzuzeigen versucht, wie bei ihrer Anwendung Rationalität „zum Zug kommt". Aber mindestens ebenso wichtig, ja noch wesentlich wichtiger ist es, diese gesetzlichen Vorschriften selbst kritisch-rational zu hinterfragen, und ebenso die in dieser Auseinandersetzung wieder vorausgesetzten Ziele und Zwecke, usw.

Die Einforderung von Rationalitätsstandards darf nicht bei der kritischen Hinterfragung von Mitteln stehenbleiben. Würde die Rationalität der Überprüfung vor den Zwecken haltmachen, blieben die entscheidenden Fragen, um die es in der rechtlich-normativen und auch in der ethisch-normativen Auseinandersetzung geht, von einer kritisch-rationalen Diskussion ausgeklammert. Mit der bloßen „Zweckrationalität" (d.i. Rationalität relativ zu vorausgesetzten Zwecken) ist uns aber in Recht und Ethik – so wichtig sie auch hier ist – zu wenig gedient. Mit gutem Grund verlangen wir, daß sich in Recht und Ethik auch die Ziele und Zwecke als vernünftig ausweisen lassen – und zwar durch rationale Argumente. Es stimmt, daß es sich hier um Diskussionen auf zwei verschiedenen Ebenen handelt; bloß deshalb, weil eine Diskussion auf einer anderen Ebene geführt werden muß, ist sie noch lange nicht überflüssig oder gar unmöglich. Kein noch so hohes und hehres Ziel kann und darf sich einer kritisch-rationalen Hinterfragung entziehen. Richtig am Standardmodell der Rationalität ist, daß jede kritisch-rationale Hinterfragung ihrerseits Ziele und Zwecke voraussetzt, die im Rahmen der jeweiligen Hinterfragung außer Streit gestellt werden müssen. Aber kein Ziel und kein Zweck ist deswegen sakrosankt und gegen rationale Kritik immun – man muß bei einer solchen Kritik bloß jeweils wieder

ein anderes Ziel bzw. einen anderen Zweck voraussetzen. Vom durchaus berechtigten Prinzip „De gustibus non est disputandum" führt keine Brücke zu „De finibus non est disputandum", auch wenn noch so viele diese „schwindelige" Brücke benützen.

Die Rationalitätsstandards dieser Diskussionen „auf höherer Ebene" bleiben genau dieselben wie bei den „niedrigeren" Diskussionen. Einzig und allein die vorausgesetzten Prinzipien werden immer allgemeiner. Außerdem wird in der rechtlichen Diskussion einmal der Punkt erreicht, bei dem wir keine weiteren positiv-rechtlichen Gesetze voraussetzen können und daher auf Prinzipien (wie z.B. eine „Grundnorm") zurückgreifen müssen, die wir nicht mehr einem positiv-rechtlichen Kodex entnehmen können. Die Logik der Argumentation wird dadurch jedoch nicht beeinträchtigt. Jedenfalls muß gelten: De gustibus non est disputandum; de finibus disputandum est.

d) Damit ist in groben Zügen abgesteckt, welche Stufe der Rationalität in Recht und Ethik erreichbar ist. Ist damit auch schon gesagt, daß diese Art der Rationalität in rechtlichen und ethischen Diskussionen auch wünschenswert ist? Daß für Recht und Ethik der höchste auf diesen Gebieten erreichbare Grad an Rationalität auch wünschenswert ist, ergibt sich aus der Aufgabe von Recht und Ethik, unser Handeln zu steuern oder ihm zumindest eine Orientierungshilfe zu bieten. Diese Aufgabe kann nämlich nur erfüllt werden, wenn für konkrete Anwendungsfälle einigermaßen klar bestimmt ist, wie sie aus der Sicht von Recht und Ethik zu beurteilen sind. Ohne Rationalität bleibt dies für die Handelnden völlig unklar und unbestimmt; Rechtssicherheit und ethische Orientierung gehen dadurch verloren und damit auch der Steuerungseffekt von Recht und Ethik. Recht und Ethik können in umso höherem Maß ihre Steuerungs- und Orientierungsaufgabe erfüllen, je höher das Ausmaß an Rationalität bei ihrer Anwendung ist.

Appendix: Die Anfänge der Imperativ- und Normenlogik, die „Logic of Satisfaction" und die „Logic of Satisfactoriness"

Die „Logic of Satisfactoriness" wurde von Anthony Kenny als Antwort auf die „Logic of Satisfaction" entwickelt, gegen die gravierende Einwände erhoben worden waren. Die „Logic of Satisfaction" entstand ihrerseits aus dem Bemühen, eine semantische Grundlage für eine Logik der Imperative und der Normsätze zu finden, obwohl diese Sätze keine Wahrheitswerte haben können. Zum Verständnis der „Logic of Satisfactoriness" ist es daher erforderlich, auf die „Wurzeln" der Imperativ- und Normenlogik zurückzublicken.

Ernst Mallys kühner Versuch, gewissermaßen aus dem Nichts ein axiomatisches System für die „Deontik" bzw. für die Logik der Sollsätze zu entwickeln (Mally (1926)), ist gründlich gescheitert: Aus seinen Axiomen ließen sich nämlich eine ganze Reihe von offenkundig absurden Konsequenzen ableiten, die er zum Teil selbst als Theoreme seines Systems angeführt und als „befremdlich" (Mally (1926), 20–25, 34–36) oder gar „paradox" (Mally (1926), 48, 56, 68) eingestuft hat. Bei einigem guten Willen hätte sich Mallys Axiomensystem durch entsprechende Eingriffe ohne weiteres „reparieren" lassen (vgl. Morscher (1998), 107 f.). Statt dessen haben die Philosophen lieber eine Zeitlang ganz ihre Finger von diesem Thema gelassen. Diese Reaktion auf Mallys Pionierwerk

wurde noch dadurch verstärkt, daß Anfang der 30er Jahre aufgrund der Arbeiten von Alfred Tarski immer klarer wurde, daß ein formales Axiomensystem allein noch keine Logik ausmacht, sondern daß eine Logik auch eine semantische Grundlage braucht. Mally hatte sich mit intuitiven Erläuterungen behelfen müssen, da damals noch keine wissenschaftliche Semantik zur Verfügung stand.

Tarski hat als erster strenge Definitionen von grundlegenden semantischen Begriffen wie den Begriffen der Erfüllung, der Wahrheit und der Falschheit aufgestellt. Mit Hilfe dieser Begriffe hat er dann auch logische Grundbegriffe wie insbesondere den Begriff der logischen Folgerung erstmals in der Geschichte der Logik streng definiert. Diese wichtigen Resultate von Tarski führten zu einer zusätzlichen Verunsicherung bei denjenigen, die sich mit der logischen Analyse von Imperativen und Normsätzen beschäftigten und nach einer Logik (inklusive einer Semantik) für solche Sätze Ausschau hielten. Nach der damals weit verbreiteten und bei kritisch eingestellten Philosophen vorherrschenden non-kognitivistischen Lehrmeinung können nämlich Imperative und Normsätze gar keine Wahrheitswerte haben oder jedenfalls nicht wahr oder falsch im gewöhnlichen Sinn dieser Wörter sein, in welchem sie nur auf Aussagesätze anwendbar sind. Aus dem Non-Kognitivismus leiteten manche Philosophen ab, daß eine Logik für Imperative und Normsätze gar nicht möglich sei, und zwar in dem strengen Sinn, daß Imperative und Normsätze überhaupt keine logischen Beziehungen eingehen können – weder untereinander noch zu anderen Sätzen (wie z.B. zu Aussagesätzen). Folgende Überlegung stand dabei Pate: Wenn ein formaler Logikkalkül korrekt ist, dann kann in diesem Kalkül ein Satz aus anderen Sätzen nur dann ableitbar sein, wenn er aus ihnen logisch folgt; der Begriff der logischen Folgerung ist aber – bei Tarski zumindest – nur für Sätze definiert, für die auch die Begriffe der Wahrheit und Falschheit (bzw. der Erfüllung) definiert sind. Wenn aber für Imperative und Normsätze die Begriffe der Wahrheit und Falschheit gar nicht definiert sind bzw. wenn diese Sätze gar nicht wahr oder falsch (im gewöhnlichen, von Tarski definierten Sinn dieser Wörter) sein können, dann sind für diese Sätze auch andere semantische Begriffe wie der Begriff der logischen Folgerung, der Verträglichkeit und Unverträglichkeit usw. nicht definiert, für deren Definition (zumindest bei Tarski) der Begriff der Wahrheit (bzw. der Erfüllung) vorausgesetzt wird. In weiterer Konsequenz können dann aber auch Ableitungen bzw. Schlüsse, die solche wahrheitswertlosen Sätze wie Imperative oder Normen als Prämissen oder Konklusion enthalten, nicht als logisch korrekt legitimiert werden, auch wenn sie gewissen Schlußregeln entsprechen.

Diesen theoretischen Einwänden und Skrupeln zum Trotz gibt es eine Vielzahl von Schlüssen, die ganz offenkundig logisch korrekt sind und dennoch Imperative bzw. Normsätze enthalten wie z.B.: Hilf Deinem Nachbarn, wenn immer er in Not ist! Dein Nachbar ist jetzt in Not; daher: Hilf jetzt Deinem Nachbarn! Jørgen Jørgensen hat diesen Zwiespalt zwischen Theorie und Praxis des imperativischen bzw. normativen Schließens als „puzzle" thematisiert (Jørgensen (1937/38), 290); Ross (1941), 55 f., gab dem Problem den Namen „Jørgensen's Dilemma", unter dem es bis heute in der Fachliteratur diskutiert wird. Da sich die erfolgreiche und offensichtlich korrekte Praxis des normativen Schließens nicht durch theoretische Bedenken „abschaffen" ließ und dafür auch ein Bedarf in verschiedenen Lebensbereichen (wie etwa in der Politik, vor allem aber auch im Rechtsleben) besteht, machte man sich auf die Suche nach einer theoretischen Begründung für die praktisch klaglos funktionierende Logik der Imperative und Normsätze.

Es war naheliegend, auf der Suche nach einer Logik für Imperative und Normsätze, die selbst keine Wahrheitswerte haben können, nach Aussagesätzen Ausschau zu halten, die in einer derartigen Verbindung mit den Imperativen und Normsätzen stehen, daß die für diese Aussagesätze zur Verfügung stehende Logik auch auf die Imperative und Normsätze „übertragen" werden kann. Ein Lösungsansatz für dieses Problem besteht darin, die Imperative bzw. Sollsätze definitorisch auf Aussagesätze zu reduzieren. Beispiele dafür sind: (1) ‚Es ist gesollt (oder geboten), daß p' (bzw. kurz: ‚$O(p)$') besagt dasselbe wie: ‚Es gibt einen Normgeber (z.B. Gott, ein Parlament etc.), der gebietet, daß p'; (2) $O(p) :\leftrightarrow$ die Mehrheit der Bevölkerung will, daß p; (3) $O(p) :\leftrightarrow \Box(\neg p \rightarrow S)$, bzw. logisch äquivalent damit, aber psychologisch vielleicht wirksamer (vgl. z.B.: ‚Geld oder Blut!'): $O(p) :\leftrightarrow \Box(p \vee S)$, wobei ‚$S$' ein Satz ist, der eine bestimmte Sanktion beschreibt. (Die Definition (3) wurde bereits von Bohnert (1945), 311, vorgeschlagen. Auf dieser Grundidee baut das formale System von Anderson (1956) auf, in dem die deontische Logik auf die alethische Modallogik zurückgeführt wird; vgl. auch Anderson (1958) und Prior (1958).)

Lösungen dieser Art sind aber naturalistisch und daher auch kognitivistisch; sie kamen daher für die Non-Kognitivisten nicht in Frage. Diese suchten infolgedessen nach einem anderen Ausweg: Den Imperativen bzw. Normsätzen werden dabei Aussagesätze zugeordnet, deren Logik auf die Imperative bzw. Normsätze „übertragen" werden kann, ohne daß diese dadurch auf die betreffenden Aussagesätze reduziert werden (wie bei der ersten Lösung). Man geht dabei von der Überlegung aus, daß jeder Imperativ und jeder Normsatz einen Inhalt hat, der durch einen Aussagesatz wiedergegeben wird; in der formalen Darstellung eines Normsatzes ‚$O(p)$' bzw. eines Imperativs ‚$!(p)$' steht die Variable ‚p' für einen beliebigen Aussagesatz, welcher dem Inhalt des Normsatzes ‚$O(p)$' bzw. des Imperativs ‚$!(p)$' entspricht und den Zustand beschreibt, der durch den Normsatz bzw. Imperativ vorgeschrieben wird. Dieser Aussagesatz beschreibt zugleich den Zustand, der besteht, wenn der Normsatz ‚$O(p)$' bzw. der Imperativ ‚$!(p)$' erfüllt (oder befolgt) wird. Der in einem Normsatz ‚$O(p)$' bzw. Imperativ ‚$!(p)$' anstelle der Variablen ‚p' stehende Aussagesatz ist somit der zu ‚$O(p)$' bzw. ‚$!(p)$' gehörige „Erfüllungssatz". Diese Überlegungen führten Walter Dubislav zu folgender Festlegung, welche für ihn die Grundlage für den logischen Umgang mit Forderungssätzen (d.s. Imperative bzw. Normsätze) bildet: „Ein Schliessen aus Forderungssätzen wird nun formal durch nachstehende Vereinbarung ermöglicht: Ein Forderungssatz F heisst ableitbar aus einem Forderungssatz E, wenn der zu F gehörende Behauptungssatz im üblichen Sinne aus dem zu E gehörenden ableitbar ist" (Dubislav (1937), 341). Unter dem zu einem Forderungssatz F gehörenden Behauptungssatz versteht Dubislav genau das, was ich zuvor den zu F gehörigen Erfüllungssatz genannt habe. Ich werde im folgenden den zu einem Imperativ bzw. Normsatz N gehörigen Erfüllungssatz mit ‚Erfüllt(N)' abkürzen, was zu lesen ist als: N ist erfüllt; zugleich werde ich diese Abkürzung aber auch autonym zur *Bezeichnung* solcher Sätze verwenden. (Für die Metasprache werde ich übrigens außer ‚$\Rightarrow$' und ‚$\Leftrightarrow$' keine eigenen logischen Symbole einführen, sondern einfach die bereits zuvor in objektsprachlichen Sätzen benützten Symbole verwenden, weil auch daraus kaum ein Mißverständnis erwachsen kann.)

Die „Vereinbarung" von Dubislav ist wohl als eine Art Definition des Begriffs der Ableitbarkeit für Imperative bzw. Normen gedacht. Zur damaligen Zeit hat man jedoch noch nicht so klar wie heute zwischen dem syntaktischen Begriff der Ableitbarkeit ($\vdash$)

und dem semantischen Begriff der logischen Folge(rung) ($\models$) unterschieden. Wenn auch Dubislav und andere Autoren, von denen hier noch die Rede sein wird, explizit von Ableitbarkeit sprachen, hatten sie dabei doch primär den semantischen Folgerungsbegriff im Auge. Ich werde daher ihre Ausführungen hier stillschweigend in die semantische Terminologie übersetzen.

Eine Logik, die auf der Vereinbarung von Dubislav oder einer ähnlichen Festlegung beruht, hat man später aus naheliegenden Gründen ‚Erfüllungslogik' genannt. Man kann daher Dubislavs Vereinbarung als Definition wiedergeben, in welcher der „erfüllungslogische" Folgerungsbegriff für Normen bzw. Imperative ($\models_E$) mit Hilfe des „normalen" Folgerungsbegriffs für Aussagesätze ($\models$) definiert wird:

$$\text{DD: } N_1 \models_E N_2 :\leftrightarrow \text{Erfüllt}(N_1) \models \text{Erfüllt}(N_2)$$

Aus DD erhalten wir für Normen der Form ‚$O(p)$' und ‚$O(q)$' den Spezialfall

$$\text{DD': } O(p) \models_E O(q) \Leftrightarrow p \models q$$

Strenggenommen hat Dubislav allerdings seine Vereinbarung nur als Wenn-dann-Satz formuliert, und in dieser Form können wir sie als Dubislavs Postulat (DP) festhalten:

$$\text{DP: Erfüllt}(N_1) \models \text{Erfüllt}(N_2) \Rightarrow N_1 \models N_2$$

bzw.

$$\text{DP': } p \models q \Rightarrow O(p) \models O(q)$$

Durch dieses Postulat wird die Verwendbarkeit des „normalen" Folgerungsbegriffs auf bestimmte Arten von Normen bzw. Imperativen (nämlich auf Normen der Form ‚$O(p)$' und auf Imperative der Form ‚$!(p)$') ausgedehnt. Für eine Verallgemeinerung dieser Idee wäre es erforderlich, den Begriff des Erfüllungssatzes für beliebige Normen und Imperative zu definieren. Außerdem werden als Vorderglied der Folgerungsrelation einer solchen Imperativ- bzw. Normenlogik von Dubislav immer nur einzelne Imperative bzw. Normsätze in Betracht gezogen, nicht jedoch ganze Satzmengen, zu denen unter Umständen neben Imperativen bzw. Normsätzen auch Aussagesätze gehören können. (Dubislav (1937), 341, definiert zwar noch einen erweiterten Ableitbarkeits- bzw. Folgerungsbegriff, bei dem die Ableitbarkeit bzw. Folgerung eines Imperativs aus einem anderen Imperativ auch von zusätzlichen Aussagesätzen abhängt, diese Aussagesätze gehören aber bei ihm nicht zur Prämissenmenge; Hofstadter & McKinsey (1939), 457, erwähnen zwar ausdrücklich die Möglichkeit, auch Aussagesätze als Prämissen zuzulassen, doch sie gehen dieser Möglichkeit nicht nach.) Ich sehe bei meiner Darstellung von solchen Verallgemeinerungen ab, da sie auch von Dubislav und den anderen Autoren nicht weiter verfolgt wurden.

Ein formales System der Imperativlogik im Geiste von Dubislavs Grundidee haben Hofstadter & McKinsey (1939) entwickelt. Obwohl 13 Jahre nach Mallys Pionierwerk entstanden, leidet dieses System jedoch immer noch an einigen der gravierenden Kinder-

krankheiten von Mallys System. So ist in diesem System ebenso wie bei Dubislav (ohne daß sich dieser allerdings auch noch damit gebrüstet hat) z.B. die folgende Schlußform gültig: $O(p), p \rightarrow q \therefore O(q)$ (Hofstadter & McKinsey (1939), 457). Außerdem betrachten die Autoren einen unhaltbaren Notstand sogar als Tugend ihres Systems: Der Imperativ- bzw. Solloperator ‚!' ist darin nämlich angeblich – übrigens ebenso wie in Mallys System – insofern überflüssig, als es zu jedem Satz mit diesem Operator einen damit deduktiv äquivalenten Satz ohne diesen Operator gibt (Hofstadter & McKinsey (1939), 453). Das Theorem, das dieser Behauptung zugrunde liegt (Hofstadter & McKinsey (1939), 452), wird allerdings nicht bewiesen und ist in der Form, in der es formuliert wird, glücklicherweise auch gar nicht beweisbar. So schlimm, wie Ross glaubt, ist es um das System denn doch nicht bestellt: Hofstadter & McKinsey (1939), 456, schließen ausdrücklich aus, was Ross (1941), 61, befürchtet, daß nämlich in ihrem System Schlüsse der Form $p \therefore O(p)$ bzw. $p \therefore !(p)$ gültig sind.

Der Begriff der Ableitbarkeit bzw. der logischen Folge wird nun aber – und das ist im vorliegenden Kontext das Entscheidende – von Hofstadter und McKinsey (1939), 452, genau gleich wie bei Dubislav definiert: „Suppose that $C_1 = !S_1$ and $C_2 = !S_2$ are provable. Then we call C_2 *derivable from* C_1, if S_2 is derivable from S_1, and C_2 a consequence of C_1, if S_2 is a consequence of S_1". Da dabei nach Hofstadter & McKinsey die Erfüllung („Satisfaction") eines Imperativs das Analogon zur Wahrheit eines Aussagesatzes bildet (Hofstadter & McKinsey (1939), 447), prägte Ross für diese Logik den Namen ‚Logic of Satisfaction' (Ross (1941), 63 ff.), wofür sich der Ausdruck ‚Erfüllungslogik', den ich schon zuvor im Zusammenhang mit Dubislav verwendet habe, als Übersetzung anbietet und auch eingebürgert hat.

Ross gelangte nun aber zur Auffassung, daß die Erfüllungslogik unhaltbar ist. Es gibt nämlich Schlüsse, die zwar im Einklang mit der Erfüllungslogik, aber ganz offenkundig im Widerspruch zu unseren Intuitionen stehen, wie z.B.: Wirf den Brief in den Briefkasten! Daher: Wirf den Brief in den Briefkasten oder verbrenn ihn! (Ross (1941), 62). Dieser Schluß, der intuitiv unkorrekt erscheint, aber von der Erfüllungslogik legitimiert wird, ist als „Ross'sches Paradoxon" in die Literatur eingegangen und hat die Form: $O(p) \therefore O(p \vee q)$. Im Lichte der späteren Entwicklung der deontischen Logik werden Schlüsse dieser Form eher als korrekt verteidigt, und ihr paradoxes Image wird damit erklärt, daß unsere Intuition von der Alltagssprache oft irregeleitet wird. Ross aber verwarf solche Schlüsse und mußte daher einen neuen Ansatz für eine Logik der Imperative und Normen finden. Er stellte aus diesem Grund der Erfüllungslogik eine Gültigkeitslogik gegenüber. Dabei wird statt der Erfüllung der Imperative bzw. Normsätze deren Gültigkeit als logischer Wert angesehen, welcher bei korrekten Schlüssen von den Prämissen auf die Konklusion übertragen wird. Unter der Gültigkeit eines Imperativs bzw. Normsatzes kann man nun aber zweierlei verstehen: seine objektive oder seine subjektive Gültigkeit. Da sich eine objektive Gültigkeit von Imperativen und Normen nicht intersubjektiv legitimieren läßt, schließt Ross diese Alternative aus (Ross (1941), 60). Die subjektive Gültigkeit eines Imperativs bzw. Normsatzes besteht hingegen immer nur in einem psychischen bzw. sozialen Faktum (nämlich z.B. im Faktum, daß ein Normgeber den Imperativ bzw. Normsatz gebietet, oder im Faktum, daß eine Person oder Personengruppe den Imperativ bzw. Normsatz akzeptiert). Auch die Interpretation der logisch korrekten Schlüsse mit Normen bzw. Imperativen als Schlüsse, welche deren subjektive Gültigkeit erhalten, ist

daher unangemessen; dies liefe nämlich nur auf eine ganz „normale" Logik für Aussagesätze hinaus, die psychische oder soziale Fakten beschreiben, was durch die imperativische bzw. normative Form der Sätze bloß verschleiert wird. Ross schlägt daher eine „Kombination" von Erfüllungslogik und subjektiver Gültigkeitslogik vor (Ross 1941), 64). Mit diesem „Kombinations"-Vorschlag von Ross ist jedoch nicht das (logische) Postulat gemeint, daß ein praktischer (d.i. ein imperativischer oder normativer) Schluß dann und nur dann logisch korrekt ist, wenn er sowohl der Erfüllungslogik als auch der subjektiven Gültigkeitslogik entspricht. (Eine Lösung dieser Art wurde erst später von Frey (1957), 465 f., und – in modifizierter Form – von Frey (1965), 376 f. vorgeschlagen; anstelle der subjektiven Gültigkeitslogik verwendet Frey dabei den sogenannten „Existenzkalkül".) Ross verbindet mit seinem Vorschlag vielmehr bloß die (empirische) Hypothese, mit einem praktischen Schluß werde der Anspruch erhoben, daß er der Erfüllungslogik gemäß und bezüglich der subjektiven Gültigkeit der Imperative „relevant" sei (Ross (1941), 64, 68).

Der Vorschlag von Ross war zwar gut gemeint, erweist sich aber für die Begründung einer Imperativ- bzw. Normenlogik als völlig wertlos: Ganz abgesehen davon, daß die Hypothese von Ross unklar formuliert ist und nicht besonders plausibel klingt, bleiben für ihn offenbar gar keine logisch korrekten Schlüsse mit Imperativen bzw. Normsätzen mehr übrig. Seine ganze Imperativ- bzw. Normenlogik schrumpft auf zwei Schlußformen zusammen, nämlich: $O(p) \therefore \neg O(\neg p)$, und: $\forall x(Fx \rightarrow O(Gx)), Fa \therefore O(Ga)$, und selbst dabei handelt es sich nur um „pseudo-logische" Schlüsse oder Enthymeme; ihre logische Korrektheit hängt nämlich von einer stillschweigend vorausgesetzten Prämisse ab, die zwar selbstverständlich erscheinen mag, für die logische Korrektheit der Schlüsse aber erforderlich ist: daß nämlich die subjektiv gültigen Imperative des betreffenden Systems keine unvereinbaren Handlungen vorschreiben (Ross (1941), 65, 68 f., 70). Damit hat aber Ross nicht eine Grundlage für eine Logik der Imperative bzw. Normsätze geschaffen, denn er kann auf diese Weise ja nicht begründen, warum gewisse Schlüsse mit Imperativen bzw. Normsätzen tatsächlich logisch *korrekt sind*; sondern er hat damit bestenfalls erklärt, warum uns solche Schlüsse intuitiv als *korrekt erscheinen*. (Zur „Ehrenrettung" von Ross könnte man höchstens anführen, daß er mit seinem Vorschlag gar nicht mehr erreichen wollte und sich mit diesem bescheidenen Resultat selbst begnügt hat.)

Auch Ross gelang es somit nicht, die Imperativ- bzw. Normenlogik, die seit Mallys Fehlschlag orientierungslos dahinschlitterte, in geordnete Bahnen zu lenken. Dies war Georg Henrik von Wright vorbehalten, der 1951 mit seinem bahnbrechenden Aufsatz „Deontic Logic" (von Wright (1951)) einen Neubeginn der Imperativ- bzw. Normenlogik einleitete und ihre kontinuierliche Weiterentwicklung in Gang setzte. Einige Jahre danach wurde von Stig Kanger der erste Ansatz einer Mögliche-Welten-Semantik entwickelt, den er 1957 in mehreren Publikationen vorstellte und auch gleich von Anfang an zur Behandlung von ethischen Normsätzen benützte (vgl. Kanger (1957)). Kurz darauf hat Saul Kripke seine Mögliche-Welten-Semantik für die Modallogik veröffentlicht (Kripke (1959a)) und auf ihre Verwendbarkeit für die deontische Logik und damit für Normsätze hingewiesen (Kripke (1959b)).

Trotz aller immer wieder vorgebrachten Bedenken gegen eine Mögliche-Welten-Semantik muß man zugestehen, daß es sich dabei um die erste Semantik für Normsätze handelt, welche den Ansprüchen Tarskis an eine wissenschaftliche Semantik genügt. Sie läßt

jedenfalls alle früheren Ansätze, eine solche Semantik für Imperative und Normsätze zu entwickeln, weit hinter sich zurück, und an ihr muß jeder Neuansatz für eine derartige Semantik gemessen werden. Im Rahmen dieser Semantik ist es nun aber auch möglich, das Dubislav-Postulat der Erfüllungslogik (DP) zu „begründen". In der Mögliche-Welten-Semantik für Normen (und Imperative) wird nämlich eine Idee der Erfüllungslogik wiederbelebt: Nach der Mögliche-Welten-Semantik ist ein Normsatz der Form ‚$O(p)$' in einer Welt w genau dann gültig, wenn der Satz, der anstelle von ‚p' steht, in jeder deontisch perfekten Alternative w' von w wahr ist, wenn also ‚$O(p)$' selbst in jeder deontisch perfekten Alternative w' von w erfüllt ist. Wenn nun aber q in jeder möglichen Welt w jeder beliebigen Interpretation wahr ist, in der auch p wahr ist, dann muß q auch in jeder möglichen Welt w' wahr sein, in der p wahr ist und die zusätzlich noch eine deontisch perfekte Alternative von w darstellt. Damit wird auch verständlich, warum so viele das Gefühl haben, daß in der Erfüllungslogik trotz ihres Scheiterns ein richtiger Kern steckt.

Die Mögliche-Welten-Semantik wurde von vielen lange nicht wahr- und auch dann oft noch nicht ernstgenommen. Das erklärt, warum die alte Erfüllungslogik und andere informelle Ansätze einer Semantik der Imperative und Normen nicht gleich ausstarben, sondern in gewissen Nischen weiterlebten und weitergepflegt wurden. (Um zumindest *einen* Beleg dafür anzuführen, sei darauf verwiesen, daß sich Kelsen noch sehr spät mit den verschiedenen normenlogischen Ansätzen, die ich hier erwähnt habe, eingehend auseinandergesetzt hat; vgl. die postume Veröffentlichung in Kelsen (1979), 154–165.) In diesem Kontext entwickelte Anthony Kenny seine „Logic of Satisfactoriness". Auch er betrachtet die Schlußform, welche dem Ross'schen Paradoxon zugrunde liegt, als logisch unzulässig (Kenny (1966), 67, 74, Kenny (1975), 73, 82). Nach einer eingehenden Auseinandersetzung mit dem Aufsatz von Alf Ross kommt Kenny zum Schluß: „I suggest that what we need in place of Ross's logic of validity is something which we may call the logic of *satisfactoriness*" (Kenny (1966), 71); bzw. später: „the logic operative in practical reasoning [is] the logic of *satisfactoriness*" (Kenny 1975), 81). Kenny benennt „seine" Logik nach der Eigenschaft, die aufgrund der Regeln dieser Logik von den Prämissen auf die Konklusion übertragen wird; da es für das Hauptwort ‚satisfactoriness' keinen passenden deutschen Ausdruck gibt, lasse ich es im folgenden unübersetzt. „Satisfactoriness" ist nach Kenny primär eine Eigenschaft von „fiats"; darunter versteht er Pläne, Projekte, Wünsche und Befehle sowie auch (ohne scharfe Unterscheidung) sprachliche Ausdrücke für sie (Kenny (1966), 68 f., Kenny (1975), 74). Der Einfachheit und der terminologischen Einheitlichkeit wegen werde ich hier die feinen Unterscheidungen Kennys zwischen den verschiedenen Arten von „fiats" vernachlässigen und seinen Begriff der „Satisfactoriness" insbesondere auf Imperative und Normsätze anwenden.

Beim Begriff „Satisfactoriness" handelt es sich, wie Kenny ausdrücklich betont, um einen relativen Begriff (Kenny (1966), 72, 73, Kenny (1975), 80, 82, 93): Pläne sind nicht an und für sich hinreichend bzw. befriedigend („satisfactory"), sondern immer nur im Hinblick auf bestimmte Zwecke. Ein Plan ist dann für einen Zweck hinreichend, wenn seine Verwirklichung zugleich die Verwirklichung der erwünschten bzw. bezweckten Zustände beinhaltet. Die Regeln der „Logic of Satisfactoriness" garantieren, daß wir im praktischen Schließen niemals von Imperativen (bzw. „fiats"), welche für einen bestimmten Zweck hinreichend sind, zu solchen gelangen, die für diesen Zweck nicht hinreichend sind; diese Regeln erhalten also die „Satisfactoriness" im gleichen Sinne, wie die Regeln

der Logik für Aussagesätze die Wahrheit der Aussagesätze und die Regeln der „Logic of Satisfaction" die Erfüllung der Imperative erhalten (Kenny (1966), 72 f., Kenny (1975), 81). Im Rahmen dieser Logik wird der Folgerungsbegriff ($\models_H$) folgendermaßen definiert:

KD1: $N_1 \models_H N_2 :\leftrightarrow$ für jede Menge G von Zwecken bzw. Wünschen gilt: Wenn N_1 hinreichend für G ist, dann ist auch N_2 hinreichend für G

Symbolsprachlich können wir diese Definition folgendermaßen wiedergeben (ich verwende dabei ‚Hinreichend(N, G)' als Abkürzung für: ‚der Plan (bzw. der ihn ausdrückende Imperativ- oder Normsatz) N ist hinreichend für eine Menge G von Wünschen'):

KD1′: $N_1 \models_H N_2 :\leftrightarrow \forall G(\text{Hinreichend}(N_1, \text{G}) \Rightarrow \text{Hinreichend}(N_2, \text{G}))$

Daß ein Plan für eine Menge von Wünschen bzw. Zwecken hinreichend ist, erläutert Kenny aber folgendermaßen: „Now a plan is satisfactory relative to a certain set of wishes, if and only if whenever the plan is satisfied every member of that set of wishes is satisfied" (Kenny (1966), 73, Kenny (1975), 81), bzw. kurz: „A is satisfactory relative to set G iff if A is satisfied every member of G is satisfied" (Kenny (1966), 73). Diese Erläuterung läßt sich in folgende Definition kleiden:

KD2: $\text{Hinreichend}(N, G) :\leftrightarrow \Box(\text{Erfüllt}(N) \Rightarrow \forall x(x \in G \Rightarrow \text{Erfüllt}(x)))$

Dieser Definition zufolge ist jede Norm hinsichtlich der Menge von Wünschen, die nur aus ihr selbst besteht, hinreichend: „Trivially, every fiat is satisfactory relative to the purpose expressed by itself" (Kenny (1966), 72, Kenny (1975), 81); formal ausgedrückt:

KT1: $\forall N(\text{Hinreichend}(N, \{N\}))$

d.h. gemäß Definition KD2:

KT1′: $\forall N\Box(\text{Erfüllt}(N) \Rightarrow \forall x(x \in \{N\} \Rightarrow \text{Erfüllt}(x)))$

Aufgrund der Definition KD2 erhalten wir ferner folgendes Theorem (Kenny (1966), 73, Kenny (1975), 81 f.):

KT2: $\forall G(\text{Hinreichend}(N_1, G) \Rightarrow \text{Hinreichend}(N_2, G)) \Leftrightarrow (\text{Erfüllt}(N_2) \Rightarrow \text{Erfüllt}(N_1))$

Beweis: $\Rightarrow$ mit Hilfe von KT1; $\Leftarrow$ (fast) rein aussagenlogisch.

Die „Logic of Satisfactoriness" ist somit ein Spiegelbild („mirror image") der „Logic of Satisfaction", wie das folgende Theorem zeigt, das sich aus KT2 durch Anwendung der Definitionen KD1′ und DD ergibt:

KT3: $N_1 \models_H N_2 \Leftrightarrow N_2 \models_E N_1$

Unter Heranziehung von DD′ erhalten wir daraus (vgl. Kenny (1966), 73, 74, Kenny (1975), 82):

KT3′: $O(p) \models_H O(q) \Leftrightarrow \models q \rightarrow p$

Wie schon bisher steht dabei ‚$\models_H$' für logische Folge in der „Logic of Satisfactoriness", ‚$\models_E$' für logische Folge in der „Logic of Satisfaction" und ‚$\models$' für den „normalen" Folgerungs- bzw. Allgemeingültigkeitsbegriff. So weit, so schlecht: Diese „Logic of Satisfactoriness" beseitigt nicht nur angebliche Paradoxien (wie das Ross'sche Paradoxon), zu deren Lösung Kenny seine „Logic of Satisfactoriness" ja eigens entwickelt hat; vielmehr legitimiert sie auf der anderen Seite auch Schlußweisen, die nicht nur (wie Kenny (1966), 74, Kenny (1975), 83, selbst zugibt) paradox erscheinen, sondern die logisch gesehen völlig unhaltbar sind. Dazu gehört vor allem die Schlußform $O(p \vee q) \therefore O(p)$, die einer Umkehrung des Ross'schen Paradoxons entspricht und in der „Logic of Satisfactoriness" gültig ist, wie sich Kenny eigens rühmt (Kenny (1966), 74, Kenny (1975), 82); oder dual dazu: $O(p) \therefore O(p \wedge q)$; und als Spezialfall davon auch: $O(p) \therefore O(q \wedge (q \rightarrow p))$ (Kenny (1966), 75, Kenny (1975), 83); und schließlich – wenn wir nicht einen „Konsistenz-Riegel" vorschieben wie Geach (1966), 76 – sogar (was Kenny verständlicherweise unerwähnt läßt): $O(p) \therefore O(p \wedge \neg p)$.

Diese paradoxen, ja zum Teil sogar absurden Konsequenzen von Kennys Theorie wurden von verschiedenen Autoren aufgezeigt und kritisiert (vgl. vor allem Hare (1971), 64 f., Gombay (1967)). Aufgrund dieser Einwände hat Kenny selbst seine „Logic of Satisfactoriness" revidiert oder zumindest modifiziert. Zunächst hatte es bei ihm geheißen: „the logic of satisfactoriness, and not the logic of satisfaction, is the principal logic of imperatives" (Kenny (1975), 73). Diese Formulierung beurteilt er einige Jahre später als mißverständlich (Kenny (1975), 85), und er erklärt nunmehr, die „Logic of Satisfactoriness" mache nur einen *Teil* des praktischen Schließens aus (Kenny (1975), 95), allerdings – trotz der Einwände von Hare – einen *zentralen* Teil (Kenny (1975), 91, 95). In diesem Zusammenhang unterstreicht Kenny besonders die „defeasibility" als charakteristischen Zug des praktischen Schließens (Kenny (1975), 92–96), auf den schon Geach hingewiesen hatte (Geach (1966), 77). Nach Kenny liegt der Grund für die „defeasibility" des praktischen Schließens darin, daß der Begriff der „satisfactoriness" relativ ist (Kenny (1975), 93), und inhaltlich bringt er sie mit der Willensfreiheit in Verbindung (Kenny (1975), 96).

Diese Erläuterungen heben jedoch nicht die absurden Konsequenzen seiner „Logic of Satisfactoriness" auf. Dafür hat allerdings Hare den Weg geebnet: Das Schließen auf notwendige Bedingungen (das der „Logic of Satisfaction" entspricht) und das Schließen auf hinreichende Bedingungen (das der „Logic of Satisfactoriness" entspricht) stehen nach Hare gleichwertig nebeneinander und ergänzen sich gegenseitig. Man müßte nun bloß – als Zuspitzung dieser Überlegung von Hare – die beiden miteinander verbinden: Ähnlich wie ja auch Ross eine Kombination von „Logic of Satisfaction" und „Logic of Validity" vorgeschlagen hat, könnte man die „Logic of Satisfaction" mit der „Logic of Satisfactoriness" kombinieren. Wir verzichten dabei auf eine Definition und begnügen uns mit der implikativen Festlegung: Eine Norm folgt logisch aus einer anderen Norm, wenn die Erfüllung der beiden Normen *gegenseitig* auseinander folgt, wenn also ihre Erfüllungssätze

miteinander logisch äquivalent sind; allerdings sind dann auch die beiden Normen miteinander logisch äquivalent. Damit ließe sich die „Logic of Satisfactoriness" auf einfache Weise „bereinigen"; sie würde dadurch nämlich auf eine harmlose These zusammenschrumpfen, die der deontischen Extensionalitätsregel entspricht, welche in den Standardsystemen der deontischen Logik gültig ist und sich im Rahmen einer Mögliche-Welten-Semantik begründen läßt:

KT4: $\models \text{Erfüllt}(N_1) \leftrightarrow \text{Erfüllt}(N_2) \Rightarrow \models N_1 \leftrightarrow N_2$

bzw.

KT4′: $\models p \leftrightarrow q \Rightarrow \models O(p) \leftrightarrow O(q)$

In meiner Fallstudie habe ich die deontische Schlußregel R2 mehrfach verwendet, welche den Übergang von ‚$O(p)$' und ‚$\Box(p \rightarrow q)$' zu ‚$O(q)$' gestattet, und ich habe diese Regel in Verbindung mit der „Logic of Satisfaction" gebracht. In Wirklichkeit handelt es sich dabei jedoch um eine Regel, die mit der „Logic of Satisfaction" und deren Postulat DP′ bloß „geistesverwandt" ist. Aus naheliegenden Gründen mußte ich aber für meine Darstellung entsprechende Adaptierungen vornehmen. Dasselbe gilt auch für die Schlußregel (oder genauer: die „Fehlschlußregel"), die den Übergang von ‚$O(p)$' und ‚$\Box(q \rightarrow p)$' zu ‚$O(q)$' gestattet und die ich der „Logic of Satisfactoriness" zugeordnet habe. Durch den Appendix sollte der historische Hintergrund dieser beiden Schlußregeln ausgeleuchtet werden. Vor allem wollte ich damit aber vor der „Logic of Satisfactoriness" warnen, die in wesentlich weniger gefinkelter Form unser Denken in praktischen Fragen häufig verführt und irreleitet. Dabei kommt es oft auch zur Verwechslung von notwendigen und hinreichenden Mitteln für einen bestimmten Zweck, deren Unterscheidung das Anliegen von Hare (1969) war. Einige solche Mißverständnisse habe ich in anderen Arbeiten aufzuklären versucht; so etwa in Hinblick auf politische Entscheidungen in Morscher (1984) und in Hinblick auf methodologische Regeln in Morscher (1985).*

Literatur

Anderson, A.R. 1956 *The Formal Analysis of Normative Systems*, New Hawen: Veröffentlicht als *Technical Report*, Nr. 2 (Contract No. SAR/Nonr-609 (16), Office of Naval Research, Group Psychology Branch). Nachdruck in N. Rescher (ed.), *The Logic of Decision and Action*, Pittsburgh: University of Pittsburgh Press, 1967, 147–213.

Anderson, A.R. 1958 „A Reduction of Deontic Logic to Alethic Modal Logic", *Mind* 67, 100–103.

* Dieser Aufsatz ist Teil eines größeren Projekts über Paradigmen des praktischen Schließens im Rahmen von Projektteil 9 (Ethik) des Spezialforschungsbereichs F012 „Theorien- und Paradigmenpluralismus in den Wissenschaften" an der Universität Salzburg.

Bohnert, H.G. 1945 „The Semiotic Status of Commands", *Philosophy of Science* 12, 302–315.

Dubislav, W. 1937 „Zur Unbegründbarkeit der Forderungssätze", *Theoria* 3, 330–342.

Foregger, E. und Fabrizy, E.E. 1999 *Strafgesetzbuch (StGB) samt ausgewählten Nebengesetzen*, 7.Aufl., Wien: Manz.

Frey, G. 1957 „Idee einer Wissenschaftslogik. Grundzüge einer Logik imperativer Sätze", *Philosophia Naturalis* 4, 434–491.

Frey, G. 1965 „Imperativ-Kalküle", in K. Ajdukiewicz (ed.), *The Foundation of Statements and Decisions. Proceedings of the International Colloquium on Methodology of Sciences held in Warsaw, 18–23 September 1961*, Warszawa: PWN – Polish Scientific Publishers, 369–383.

Geach, P.T. 1966 „Dr. Kenny on Practical Inference", *Analysis* 26, Nr.3, 76–79. Nachdruck in P.T. Geach, *Logic Matters*, Oxford: Blackwell, 1972, 285–288.

Gombay, A. 1967 „What *is* Imperative Inference?", *Analysis* 27, Nr.5, 145–152.

Hare, R.M. 1969 „Practical Inferences", in V.Kruse (ed.), *Festskrift til Alf Ross*, Copenhagen: Juristforbundets Forlag. Nachdruck in R.M. Hare, *Practical Inferences*, London-Basingstoke: Macmillan, 1971, 59–73.

Hart, H.L.A. 1961 *The Concept of Law*, Oxford: Clarendon Press. Nachdruck (with corrections) 1972.

Hofstadter, A. und McKinsey, J.C.C. 1939 „On the Logic of Imperatives", *Philosophy of Science* 6, 446–457.

Jørgensen, J. 1937/38 „Imperatives and Logic", *Erkenntnis* 7, 288–296.

Jørgensen, J. 1938 „Imperativer og Logik", *Theoria* 4, 183–190.

Kanger, St. 1957 *New Foundations for Ethical Theory, Part 1*, Stockholm: Privat vervielfältigt. Nachdruck in R. Hilpinen (ed.), *Deontic Logic: Introductory and Systematic Readings*, Dordrecht: Reidel, 1971, 36–58.

Kelsen, H. 1968 „Zur Frage des praktischen Syllogismus", *Neues Forum* 15, Nr.173, 333–334.

Kelsen, H. 1979 *Allgemeine Theorie der Normen*, hg. von K. Ringhofer und R. Walter, Wien: Manz.

Kenny, A. J. 1966 „Practical Inference", *Analysis* 26, Nr.3, 65–75.

Kenny, A. 1975 *Will, Freedom and Power*, Oxford: Blackwell.

Kripke, S. A. 1959a „A Completeness Theorem in Modal Logic", *The Journal of Symbolic Logic* 24, 1–14.

Kripke, S. A. 1959b „Semantical Analysis of Modal Logic" (Abstract), *The Journal of Symbolic Logic* 24, 323–324.

Mally, E. 1926 *Grundgesetze des Sollens. Elemente der Logik des Willens*, Graz: Leuschner & Lubensky. Nachdruck in E. Mally, *Logische Schriften. Großes Logikfragment – Grundgesetze des Sollens*, hg. von K. Wolf und P. Weingartner, Dordrecht: Reidel, 1971, 227–324.

Morscher, E. 1973 „Philosophische Begründung von Rechtsnormen", in H. Köchler (ed.), *Philosophie und Politik: Dokumentation eines interdisziplinären Seminars*, Innsbruck: Arbeitsgemeinschaft für Wissenschaft und Politik an der Universität Innsbruck, 31–46.

Morscher, E. 1974 „Das Basisproblem in der Theologie", in E. Weinzierl (ed.), *Der Modernismus: Beiträge zu seiner Erforschung*, Graz-Wien-Köln: Styria, 331–368.

Morscher, E. 1978 „Wittgenstein's View on Ethics", in E. Leinfellner, W. Leinfellner, H. Berghel und A. Hübner (eds.), *Wittgenstein and his Impact on Contemporary Thought: Proceedings of the Second International Wittgenstein Symposium*, Wien: Hölder- Pichler-Tempsky, 494–498.

Morscher, E. 1981 „Zur ‚Verankerung' der Ethik", in E. Morscher, O. Neumaier und G. Zecha (eds.), *Philosophie als Wissenschaft/Essays in Scientific Philosophy*, Bad Reichenhall: Comes, 429–446.

Morscher, E. 1982 „Sind Moralnormen wissenschaftlich überprüfbar und begründbar?", in J. Seifert, F. Wenisch und E. Morscher (eds.), *Vom Wahren und Guten*, Salzburg: St. Peter, 102–106 und 111–116.

Morscher, E. 1984 „Ethik und politische Entscheidung", in P. Lüftenegger (ed.), *Philosophie und Gesellschaft*, Wien: Institut für Wissenschaft und Kunst, 37–47.

Morscher, E. 1985 „Wissenschaftliche Werte und methodologische Normen: Präskriptive versus deskriptive Wissenschaftstheorie", in R.P. Born und J. Marschner (eds.), *Philosophie – Wissenschaft – Politik*, Wien-New York: Springer, 181–195.

Morscher, E. 1998 „Mallys Axiomensystem für die deontische Logik: Rekonstruktion und kritische Würdigung", *ProPhil* 2, 81–165.

Nute, D. (ed.) 1997 *Defeasible Deontic Logic*, Dordrecht-Boston-London: Kluwer.

Prior, A. 1958 „Escapism: The Logical Basis of Ethics", in A.I. Melden (ed.), *Essays in Moral Philosophy*, Seattle: University of Washington Press, 135–146.

Ross, A. 1941 „Imperatives and Logic", *Theoria* 7, 53–71. Nachdruck in *Philosophy of Science* 11, 1944, 30–46.

Wright, G.H. von 1951 „Deontic Logic", *Mind* 60, 1–15.

The Invention of Western Reason

PHILIPPE NEMO

Introduction

Reason appears to be a hallmark of our modern, democratic, liberal, critical, and scientific civilisation. Most of us in the West are proud that our modern era has heralded the reign of reason in science, law, politics, and economics. But somehow, our "rational" society contains strange inconsistencies. First, we have witnessed the practical failure of what Hayek calls the Cartesian, Saint-Simonian "rationalist constructivism", and of Hegelian philosophy, which claimed to be the "absolute knowledge", and the ultimate rational explanation of reality. These two philosophical trends might well have led to the totalitarianisms of 20th-century systems which evidently indicate a decline of reason. Secondly, it seems that Western civilisation achieved its greatest success only by accepting a release of reason, through such concepts as democracy, pluralism, unrestricted freedom of criticism in the sciences, and economic freedom, which all seem to have been designed to cope with the limits of our reason. They are the various devices which social experience has found and proved efficient to circumvene to a certain degree our ignorance in political, scientific and economic fields. But of this non-fulfilment, be it a fault or failure of reason, there is a third and remarkable manifestation in the very birth of the so-called modern civilisation which claims to be the rational civilisation par excellence.

By "modern" civilisation today, we often mean the West – "the West" being an historical artefact designed between the 11th and 13th centuries in Western Europe. I will argue that the very process of its birth, insofar as we can analyse it retrospectively, is not fully rational. It seems to appear more as a "miracle", in that the process by which reason was attained seems to be itself highly irrational. My purpose is this article is to describe a process which occurred during three centuries of the Middle Ages, and which resulted in the "invention of Western Reason". But this purpose is not, strictly speaking, an historical one. Such musings are beyond the scope of this short paper. I have already elaborated a detailed understanding of reason's development in a recent book and a paper (See Nemo: 1998 and 1999). My purpose is rather to bring to light the irrational face of this history, and to focus more precisely on the "spiritual" element of it.

Before proceeding further, it will be useful to define some of the vocabulary I will be using. "Spiritual" refers to "spirit" and "mind", but these words appear (to me) to be ambiguous in the English language, while the German "Geist" is too metaphysical for what I have in mind. The French "esprit", I believe, would be a more accurate term, but I cannot use it here. Whatever the best term, by "spiritual" elements, I mean something which undoubtedly *acts* in history, although it is not *visible*; something which organises scattered historical data (such as ideas, institutions, and facts), but which we cannot see directly, at least beforehand. We can see the nascent structures, we can discover and verify that the data is becoming coherent and meaningful, but we cannot directly see the

B. Brogaard, B. Smith (Hg.), *Rationality and Irrationality / Rationalität und Irrationalität*, S. 224–241.

organising centre by which this organising process is achieved. This invisible organising process is what I mean by "esprit".

Bergson has explained this very convincingly in the case of philosophical and artistic works. He argues that when a philosopher or an artist begins to work, he does not know exactly what his work will be, nor how his works will evolve in later years. Nevertheless, all of his works are related, one to the other, eventually achieving a unified design, or singular character. These common features and this unified design will appear only at the end, in retrospect: when you read all the books of a philosopher, you discover that all of them were pursuing the same purpose. However, no one could have predicted this beforehand, not even the author. Authors and artists are secretly guided, Bergson argues, by an "intuition" – all that he was making was secretly organised by this invisible principle. Intuition might possibly be revealed at the end, but only by the traces on what the artist or the philosopher has left behind. It is fundamental to note that this principle, although invisible, is definitely *real*. It can – even though *a posteriori* – be described, and is therefore, not an imaginary, "mystical" being, but a positive reality which does work in the real world and does produce concrete effects.

I wish to understand within the confines of this article how Western reason was designed, at a certain historical juncture in the Middle Ages, by such a "spiritual" process. As I would posit, Western "Reason" is not the fruit of "reason" – at least, if we want to say it is, we must widen the sense of this term. I have a further, more general purpose. I will offer some other examples of "turning points" in History which seem to be explicable by the same paradoxical logic. I will end by drawing some conclusions on "Reason in History", if I dare use this Hegelian expression in a non-Hegelian, and even anti-Hegelian sense. The failure of historicism (so convincingly demonstrated by Karl Popper) must not make us believe that, now, no philosophy of history at all is possible.

1. The Papal Revolution

The process which occurred between 11th and 13th centuries in Europe began with what is currently known as the "Gregorian Reform". Contrary to the work of other historians, the American historian Harold J. Berman (1970) has termed the latter event the "Papal Revolution", firstly, because it was achieved not only by Pope Gregory VII himself, but by other popes, as well as Roman clerics and intellectuals, both before and after Gregory. Secondly, he has employed this term because it was not only a "reform" – a limited, piecemeal change – but a "revolution", a complete change, a general re-organisation of knowledge, values, laws and institutions, which resulted in the birth of a new, original civilisation – the West.

This revolution was the result of a crisis in the Church at the end of the High Middle Ages, in the 11th century. Church authority was restricted by secular powers, which often dominated it, preventing it from playing its own leading, "spiritual" role. There had been some efforts to cope with these difficulties, for instance, in the 10th century, the Clunisian Reform, which resulted in the creation of many independent monasteries throughout Europe. But the desired changes were achieved by the popes themselves, especially Gregory VII and his successors. Gregory declared, in his famous "*Dictatus papæ*" (1076), that the

Pope had the "*plenitudo potestatis*", both over the Church and indirectly, over the secular kingdoms. It was the first model of an "absolute monarchy" in Europe. He decreed that priests would no longer be married, and would so constitute an independent, solitary corps, the riches of which would no longer be scattered. He decided that bishops, abbots, and clerics would be appointed by spiritual authorities. Thus, the "*libertas Ecclesiæ*" would be recovered.

Gregory VII next decided that Roman law, which had been almost completely forgotten in Europe, would be studied again. The adoption of Roman Law resulted in the creation of the first European University, established by Irnerius in Bologna – designed in part to be a vehicle for the spread of Roman Law. Gregory's purpose was to give technical models for the new canonic law, created by the Papacy, which would collect old ecclesiastical canons, and decree new rules (the "decretals"), according to their new absolute monarchical power. Large ecumenical Councils (such as Lateran, Lyon, etc.) were summoned, and these created a new universal legislation which organised Christian society. Such changes were very new, and perhaps extremely controversial for the time. While "heresies" (among which Judaism was included) were severely attacked, the main aim of such reforms was to rationally organise economic, social, and even private lives throughout the Christian world. Soon a new "*Corpus juris canonici*", the "Decree" of Gratian (1140), was elaborated, and was incessantly updated and improved in the following centuries. Generally, a coherent system of law was being developed at this time, and it became increasingly common to use legal proceedings to decide disputes, instead of violence. A more structured, ordered society was slowly being constructed.

Many universities were established throughout Europe at this time, often by the popes, as a means of reducing local ecclesiastical and royal powers. The popes also took initiative in creating new universal monastic orders, notably the mendicant orders, such as the Dominicans and Franciscans. These were not to be contemplative, but active orders, through which Rome could control a greater part of Christian temporal life, and even political life. At that time, too, science was developed in the universities. While it is true that the first important results, really new by comparison with Greco-Roman science, would be reached only in Modern times, medieval scolasticism paved the way. Scientific methods were standardised during this time, and Greco-Roman science was studied afresh. The recovery of many manuscripts in Spanish, Byzantine, and Arab libraries contributed greatly to knowledge. Science and technology also benefited from the development of commerce and navigation.

Taking both the Papal monarchy and the old Roman Empire as models, deriving benefit from the study of the Roman public law in the universities, and using these institutions to train a new class of civil servants, states began a long, but ultimately victorious fight against feudalism. They began to centralise their administrations, to collect non-feudal taxes, to hold strong, permanent armies, and to judge on appeal from every local court. Most importantly, states began to gather together the feudal lords in the capital town of the king, thus increasing the prerogatives of royal control.

As all of these great changes were taking place in Western society, there was a tremendous increase in the various powers of the West. Between the 11th and 13th centuries, we can see a remarkable increase in population size, the growth of the existing towns and the birth of many new ones. There was also notable economic growth. These developments resulted in new geopolitical powers, as evidenced by the Crusades, the "*Reconquista*" in Spain, the "*Drang nach Osten*" movement of the Germans towards the Slavic countries, the Christianisation of Scandinavia, as well as Marco Polo's journeys and the improvements of navigation, which allowed European explorers to venture well outside the Mediterranean Sea.

One might indeed ask: what relationship existed between the success of what appears now as a new, original civilisation (the West) – and this "Papal Revolution"?

I would reply firstly that men of the West could, at that time, both *know the world* and *co-operate to act on it* better than they ever could before, and secondly that it is the "Papal Revolution" which made possible this new use of reason – which effectively "invented Western Reason".

2. From All-or-Nothing to Measure

Berman sheds light on the *theological* roots of these papal initiatives. I say "theological", but I could equally say "philosophical" or "spiritual". In any case, these initiatives marked a change within the intellectual world, within the world of ideas, the internal world, not a contingent, external change in material things. More precisely, it was a change in the "vision of the world", and, it seems to have been unpredictable – a "prophetic" change. A miracle perhaps? This is a question which will require elaboration as this paper progresses. It is important first of all, however, to understand the stages by which a new vision of the world was achieved.

First, there was the will of the Papacy to "christianise the world", in order to make it able to attain its eschatological ends. When leaving the world, Christ promised his swift return, which would herald the achievement of the messianic and eschatological prophecies of the Old Testament. During the first few centuries of the Roman Church's existence, adherents believed that Christ's coming was imminent. However, after one thousand years had passed, and nothing had transpired, many began to question the established Church teachings. Perhaps, argued some theologians, Christ was not going to return to earth because the world was not worth His coming. Man had let the world become worse and worse, and it was obvious that, from the conversion of the Roman Empire onwards, while there were Christians in the world, the world itself was not a Christian one. Christians had lived in a sinning world, praying to escape it, but alas – they had not yet attempted to transform it. In the High Middle Ages, the most admired persons in the Christian world were the monks, precisely because they lived "outside" the world, and therefore seemed to have a foot on the ladder which ascended to Heaven. The problem was that by abstaining from transforming the world, man had let it become more and more filled with sin. Now, theologians further reasoned, the situation had become so bleak that Christ could not dwell within it.

It seemed now that the time was ripe for men to transform the world for the better, so that Christ would change his mind, come back to earth, and bring about the long awaited redemption. This was the true meaning of the "*Dictatus papæ*", and of all further measures of the Gregorian Reform. If the Pope needed to wield absolute power, if the Church needed to be free from secular control and secular society, it was because they needed to

have the power to act on the world in order to transform it. The "*libertas Ecclesiæ*" was required if the Church was to be a spiritual power superior to the temporal powers, a situation comparable to the power of prophets over kings in Old Testament. If it was necessary to have the right to change the law, to create a new canonic law, and to expand the vigour of canonic law over any customary and secular law, it was because Christians had to instigate a *revolution.*

In traditional societies, the law is fixed and is above the will of man, and any ruler daring to change the old customs would have been guilty of sacrilege. However, theologians argued that law was not superior to the will of God, as God had created a human nature that was fundamentally good, and good divine laws. Unfortunately, the sin of man had destroyed this good nature, and in no city were there laws which could be said to be equal to the true, divine "natural law". So, by making new Christian laws (the papal decretals, or the canons of the great ecumenical Councils) the Church was improving the world, making it closer to what had existed during the lost Paradise and to what would exist again at the end of times. The Church argued that it not only had a right to do this, but further – that it was bound by a most sacred duty – no matter how strongly any temporal power might resist its authority. As far as Church officials were concerned, it was legitimate to undertake drastic changes in society for the purpose of hastening the Parousia of Christ. During this period, there is little doubt that Christianity was essentially revolutionary.

Nevertheless, one cannot embark towards a remote goal without being convinced that it is at least *possible* to reach it. A serious obstacle was presented by traditional Augustinian thought, which advanced the view that human nature had been completely destroyed by sin, and therefore, no human will could ever bring about his own salvation, this being something which could only be achieved by the grace of God. This obstacle was soon overcome by several important theological changes, a good example of which was the new doctrine of *atonement* introduced by Saint Anselm, and the invention of *Purgatory*.

3. The Anselmian Doctrine of Atonement

Before discussing the work of Saint Anselm, it is important to review some of the elements of the old Augustinian doctrine that traditionally held sway. As Augustine posited, after the original sin, man deserved nothing but death, and this fault could not be compensated by human works, because the fault was infinite, whereas any human work was finite. While it was true that God could save man by his grace, nobody knew who was to be saved and who was not, and there was nothing man could do in order to change this eternal decree. Human action had no value – no good action could save, and no bad action could definitely prevent someone from being saved. The only practicable solution to this dilemma was to abstain from acting altogether, a solution wholly favoured by monastic orders, who isolated themselves from the rest of the world, and refused to act within it. Salvation could be obtained, if it were possible, only by supernatural means – through prayers, pilgrimages, or the worship of relics. Reason was not required in the magical, enchanted world of the High Middle Ages.

Saint Anselm changed the theological vision which had previously justified this attitude. Anselm was an Italian monk who joined the abbey of Bec in Normandy, and studied

as a pupil of the famous theologian Lanfranc for many years, finally becoming abbot. When William, Duke of Normandy, became the Conqueror of England in 1066, he appointed Lanfranc and Anselm, successively, to the Archbishopric of Canterbury. Through his writings, and his role as Archbishop of Canterbury, Anselm proved to be both a great thinker and, in the full sense of the term, an actor in the "Papal Revolution". He was even, I would add, a member of the revolutionary party (like so many other famous thinkers and even mystics of the time, such as Humbert of Moyenmoutiers, Saint Peter Damian, and Saint Bruno, among many others).

Saint Anselm dealt primarily with the question of sin and atonement in his two works "*De incarnatione Verbi*" and "*Cur Deus homo?*", works which directly challenged the accepted traditional doctrines of Saint Augustine. To briefly summarise Anselm's challenge to Augustinian theology – it was true that the original sin was infinite, and it was also true that it could be overcome only by an infinite merit, which no man could possibly acquire. Nevertheless, the answer lay in Christ, who was a man like no other, a man who was totally innocent, without sin, but who nonetheless suffered a horrible death. By suffering under such a totally unjust penalty, which was not the price of a sin, he won an infinite merit – a "treasure of surerogatory (*supererogatorii*) merits" – which was now available, so to speak, to redeem mankind of their sins. Salvation was no longer at stake, since the grace of God had been given, and mankind *was* saved from original sin.

While this theological reappraisal could deliver mankind from original, general sin, it was not enough, nevertheless, to save man from particular sin. In effect, man was not only guilty of original sin, but of what theologians term "actual" sins – the sins for which the individual is responsible. Fortunately, as man was a finite being, his actual sins were also, by definition, finite, and for this reason, they could be repurchased by finite compensations – concrete human good works. Anselm's theology can be understood with reference to the following balance sheet :

BALANCE SHEET

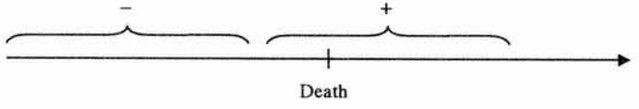
PASSIVE		ACTIVE
Original sin, actual sins		Christ's sacrifice, good works
$-\infty \quad -a-b-c\ldots$	=	$+\infty \quad +a+b+c\ldots$

This shift in the doctrine of atonement had tremendous importance for moral life. Now, human action recovers its *meaning*, since now, any concrete human action *counts in the balance*. Whatever one does, good or bad, *does matter*, because it plays an irreplaceable role in one's own personal salvation. It is up to the individual to be saved, at least up to a point. Anselm, certainly, does not negate the role of grace, since grace is necessary for one's conversion to good works. Nevertheless, while grace exists, it's really up to the individual to act well, and if he doesn't, he will not be saved by a pure miracle without his own participation in the process. In fact, Anselm's work opened the path towards what Thomas Aquinas would later posit – that grace does not act by substituting for human na-

ture, but, on the contrary, acts by restoring it, so that man can act willingly and choose freely to do good. The theological debate over whether or not grace was "sufficient", would become a Byzantine question in theology over the next few centuries. Let us remember that, with the new Anselmian theology, human action acquired a definite *value* in the face of God.

4. Purgatory

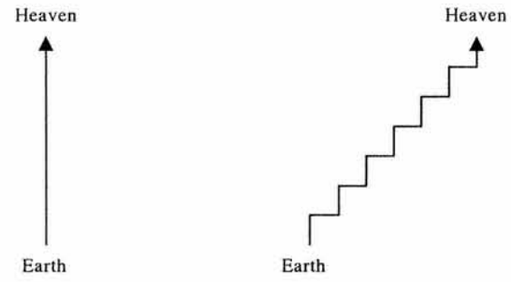
While Anselm's theories made sense to a point, there were other questions which remained unresolved. To paraphrase one of the key issues still open to theological speculation, "I will be saved if the balance is positive at the end of my life. But if I begin to do good works too late after having sinned for a very long time, then it will be difficult, and perhaps impossible, to get a positive balance when I die. In such a situation, is it worth beginning to act well?" While some could answer yes, and others no, human action would once again lose its meaning. It is not by chance that, precisely at that time (11th–12th centuries), theologians invented "Purgatory", a time after death, during which the sinner would be expected to finish his good works, in order to redeem himself.



With Purgatory, human actions once again had meaning. It was worth performing good works, even very late in one's life, even one day or one hour before death, because, even though it may not be enough to pay the entire debt, and one's deed might still hang negatively in the balance, the remainder of the debt, theologians argued, could be paid in Purgatory (the prayers of the living would also help).

5. Salvation: A Human Enterprise

With the advent of this new theology, any deeds, no matter how impressive or how insignificant, can be allotted their full value – it is now entirely up to the individual to bring about his own redemption. Human *responsibility* now becomes imperative, as Man begins to occupy centre-stage. The path towards heaven is no longer seen to be a *vertical* path, which God alone, with his magic and incomprehensible grace, or men relying on this magic alone, can ascend. Rather, the path to heaven now becomes a slant one, a series of steps, a visible way, by which man can, through rational representation, progress towards the absolute.



It is not by chance that the representation of Christ as a *mediator* is enhanced at the same time. Christ is God *and* Man. But in the High Middle Ages, as well as in orthodox Christianity until now, Christ was represented primarily in his God-like state. Even when he was depicted in painting or sculpture on the cross, Christ's triumphant face was emphasised, not his corporeal frame. Let us remember the Orthodox icons: they are representations of Godly persons, Christ and the Virgin Mary for example, but it is Christ, more than Jesus, who is painted, with his halo and infinitely serene face. He is depicted as having "risen from the dead", but without having actually died, only with an unimpaired glory. Even this minimalist representation was a controversial point in orthodoxy. By contrast, Western art began at that time to represent Christ as a suffering man, with his emaciated, injured and bloody body, and this style of representation became prevalent in all Western medieval and modern Christianity since that time. Such art emphasised the humanity of Christ, stressing that man could in fact *imitate* Him. Adopting Christ as a model was therefore not beyond human powers. From this time forward, *imitatio Christi* became the moral program of Western Christians. On the way towards Heaven, man was no longer alone, but was helped by Somebody who was like man and consequently knew and understood him. Christ showed man the straight and narrow path by which He has passed, and by which man too could travel. The ascent to salvation was no longer a question of pure grace; rather, it became, at least in part, a *human enterprise*.

The main feature of this enterprise which is important to understand was that it became a *rational* one. Salvation was no longer an "all-or-nothing" issue, but one in which man had to *measure* and *make use of his reason*.

First, he would have to calculate his own salvation, to balance evil acts by equivalent good works. Secondly, he would now have to employ his reason to achieve these good works themselves. In effect, what are "good" works? These are actions which lessen the sufferings of men, feed the hungry, support the needy, heal the sick, and generally speaking increase love and diminish evil in the world. In short, no action can be considered good which does not somehow *transform the world* for the better. But by transforming the world, man implies that he *knew* it, and will *co-operate* peacefully and efficiently within it. This therefore implies the use of reason, both in *science* and in *law* or *politics*. The use of reason now became a religious duty, which conflicted with the old duties to pray and to worship God (even though it did not substitute entirely for them). While the use of reason

was once little more than an earthly concern, and often a sinful one, it now became a moral duty *par excellence*. For it was commanded by God Himself, and was now deemed to be a path to Heaven.

6. The Grand Inquisitor

This shift in western thinking engendered a crucial misunderstanding between the Eastern and Western Churches. Eastern Christian theologians posited that when Rome decided to use reason, it effectively renounced its quest for salvation. There is a famous section in Fyodor Dostoyevski's *Brothers Karamazov*, entitled the "Legend of the Grand Inquisitor", which deals with this misunderstanding. In these pages, Dostoyevski contrasts the Roman Catholic Grand Inquisitor with Jesus. The latter is obligingly depicted as a genuine Orthodox. He is pure love and heroism, while the former is only cynical, representing the Politician *par excellence*. This politically-minded, prudent Roman Archbishop figures as a representative of Western reason, which Dostoyevski posits to be basically materialist, vulgar, and deprived of "soul".

The scene is magnificent: Jesus, who has come back to 16th-century Spain and has been recognised by the crowds who have begun to worship Him, is harshly treated, and imprisoned by the soldiers of the Archbishop. In the dead of night, the Archbishop visits Jesus in his prison cell and converses a long time with him. He explains to Him that He should not have come back on earth, because He demands too much of mankind. Mankind, the Archbishop argues, is not capable of the heroic virtues which Jesus has shown by resisting the temptations of the Devil in the desert. The Grand Inquisitor and all the Roman Church have understood that the people want only to eat and to enjoy their prosaic earthly life, even though they have to buy this by being enslaved by blind devotion to the iron hand. They are neither capable nor desirous of love and freedom. The Roman Catholic elites have accepted – in their own interest, perhaps, but what does it matter ? – to dominate these poor, miserable creatures. Deprived of any ideals, the Church has undertaken to organise the earthly lives of its subjects – making the horde entirely subservient to the religious and secular powers. By coming back unexpectedly, Jesus has interrupted this smooth reign and, by bringing transcendent ideals once again to earth, will deprive the people of their hard-earned, precarious happiness. The Grand Inquisitor shall not allow that. Consequently, Jesus will be burned at the stake the next morning.

This is, I believe, a tragic misunderstanding. Substituting a step-by-step path for a vertical one does not imply that you have renounced the goal of getting to the top – organising earthly life does not mean that you no longer believe in Heaven. On the contrary, Western theologians argued that you deserve to go to Heaven only if you have been able to improve the world by employing human nature, which was restored by the grace of God. Christians can only attain the promised of salvation if God and man *together* strive towards it. In addition – and this is the main point – it is *only* if you do work towards going to Heaven that you may "organise", in the sense of "improving", earthly life. The Dostoyevskian scene summarises, even today, I think, the gap between Eastern Europe (particularly the Russians) and the West. Since Peter the Great, the Russians have realised that they are behind the times, and they still have not understood why. They have not

advanced as far as the West, only because they have not attached the same transcendent value to human responsibility, human action, human powers.

Actually, it was the Papal Revolution – the reaffirming of the Christian eschatological goals and the emphasis put on human responsibility – which made the modern world possible. This revolution determined the spectacular development of science and law in the West from the 11th to the 13th centuries, laying the basis of Western Reason in the present day.

7. The West as a scientific and legal civilization

In effect, since salvation is no longer an all-or-nothing issue, but a measure issue, men of that time realized that they needed *instruments for measure*. Science and law are such instruments. Science knows the world, and says what is possible and impossible for men to do in the world. Law makes possible that men cooperate peacefully and efficiently. Both are tools of measure, produced by reason.

The available legal instrument was Roman law. The available instrument for science was Greek science.

a) *Roman law*. This had been almost completely forgotten in the West since Charlemagne and even before. Pope Gregory VII took the initiative in studying it once again. In Bologna, a town owned by a Pope's vassal, Princess Mathild, the first university of law (in fact, the first Western university) was established by Irnerius about 1070 (This fact is sometimes forgotten or underestimated, because later Roman law, which is a non-Christian, natural or secular law, became, together with Aristotelian political theory, a weapon against the Church and especially the Papacy ; but we must not apply to 11th century what was true only later, in 14th century, at the time of Marsilia of Padova or William of Occam). Let us recall that Roman law is really an instrument for measure, a rational tool. It aims at *jus suum cuique tribuere* ("giving everybody his own"), that is to say, to distinguish the properties, the "mine" and the "yours". And to recognize them after many changes, purchases, sales, marriage, heritage, creation or dissolution of companies. So it makes possible an easy and efficient cooperation between men, even when they achieve complex jobs implying that many independant persons work together without conflicting. Roman law was used as a technical model for the new canonic law, which was developed very much at this time both by the popes and the new ecumenical Councils, and which was collected in the famous *Decree* of Gratian (1140), as we have seen.

b) Greek science. It was also at that time that Greek science was studied once again. After the first Faculties of Law, Faculties of Arts were created, for instance in Paris at the end of the 12th century. "Arts" meant "liberal arts", the seven sciences, *trivium* and *quadrivium*, which had been established in Greek and Roman schools. True, these sciences had not been completely forgotten; they had been studied continuously in the monastic and episcopal schools. But, as they were now studied in universities outside the Church, they were beginning a life of their own. It is well known that these secular sciences, using essentially reason as opposed to revelation, were not accepted in universities without reluctance. Saint Albert the Great and Saint Thomas Aquinas had to fight fiercely in order to impose Aristotle. But this was achieved, mainly by the new papal troops (Albert and

Thomas were Dominicans).

It is absolutely necessary to understand that these rational instruments, Roman law and Greek science, existed already. For instance, the manuscript of *Corpus juris civilis*, the great collection of Roman law made by Emperor Justinian in the 6th century, had never been lost in the West (copies apparently existed everywhere). Similarly, although many new sources in Greek philosophy were found in the Arabic libraries of Spain (due to the *Reconquista* and to the Crusades), and in the Byzantine libraries, actually many other texts had been continuously copied in the monasteries and had been available for a long time. So the revival of law and science must not be construed as a contingent fact, the result of some fortuitous rediscovery of texts. The new fact was that texts which had been there for a long time found some use *now*, becoming meaningful *now*. Before the Papal Revolution, Westerners had been sleeping upon these old texts, somehow like the Arabs upon oil before the 19th century, because they had no longer, or not yet, any idea of what use they could be. If you think that you will be saved or condemned only by grace, and that human action has no value, you simply do not need to calculate the value of your actions. Accordingly you do not need a tool such as Roman law which makes subtle distinctions between evil and less evil, good and less good acts and allows for their systematic classification. Then, if you stumble upon a manuscript of the Justinian Code, you will ignore it, especially if it has become too difficult to interpret, being written in an old-fashioned, obscure language. It is only if you absolutely need to be guided in your collaboration with men, if you want to effectively interact with them while refraining from sin as much as you can, and if, then, measure has become a *vital* issue, that you are ready to make the appropriate efforts to unravel the mysteries of the *Corpus*, as Irnerius did in Bologna.

8. Progressive millennarism *versus* violent millenarism

Since the Papal Revolution, the West achieved great civilizational progress which resulted in the invention of the modern world, our democratic, liberal, scientific society. The West could do that because it kept two closely linked ideas, the moral duty to aim at the eschatological ends of mankind – to improve the world unto its final salvation – and to use the natural powers of man, reason and justice, rather than being content with waiting for a supernatural intervention of God. Hence the development of sciences in universities, the revival of the state, the development of law, and more precisely: the use of law as an instrument for transforming society. Indeed law in many ways brought about the birth of "politics" in the modern sense.

Later, at the time of the Enlightenment, the feeling that a continuous improvement of knowledge and of administering justice is possible on earth seemed to be confirmed by historical events, resulting in the secular ideology of "progress". But there is no doubt that, originally, this secular ideology was essentially religious. It was the *millennarist* idea, conveyed in Jewish apocalyptic literature and particularly in St John's *Book of Revelation*. It is the idea that, towards the end of earthly times, even before the beginning of the actual kingdom of God, there will be one thousand years (a "millennium") of happiness on Earth. This millennarist ideal is shared by Christianity and Biblical religions in

general. But the Papal Revolution, by enhancing human reason and human responsibility, separated it from another version, what we might call violent, revolutionary millennarism. Actually, at the time of Theodose, when the Roman Empire became officially Christian, both the Empire and the Church began to feel suspicious about millennarism, which seemed to them dangerously revolutionary. Millennarism was not entirely rejected, but such theologians as Origen or Saint Augustine explained that it had to be construed in a symbolic way. In fact, with the resurrection of Christ, the millennium had *already* begun. "We can and must long for a better time", the argument went, "but this time will be in Heaven. We are not supposed to act for it on Earth, and even less resort to violence."

Around the time of the Papal Revolution, millennarism was revived, notably by the works of Joachim of Flore, then by the radical wing of the Franciscans, and above all by popular demonstrations and riots which took place at the time of the Crusades, and which could spontaneously develop at times of epidemic or famine, causing restlessness among poor people in towns, especially in Northern France, Belgium and Germany. Later, new doctrines were invented by such radical sects as the "Brothers of the Free Spirit", Bohemian Hussites and Taborites, Anabaptists, and Thomas Müntzer's troops of the German War of Peasants, etc.

So the West was faced with two versions of millennarism, and a momentous choice of which to adhere to.

1) Irrational, superstitious men want to hasten the coming of the millenium by violent means. Violence is justified, in any of the theories cited above, by the same core convictions. Evil is in the heart of certain "wicked" men. These wicked can be *foreigners* (Jews, Muslims ...), more often they are *rich people* whom the poor envy (priests, nobles, merchants ...). Fortunately, there is a circle of happy few, the "saints". They have no evil in their own heart. God loves them and will help them. From this perspective, the only problem is to overcome the fears of the crowds. The "saints" have to convince them to act quickly, to *kill all the wicked*. Certainly, the armies will be unequal. But God's armies will help the saints at the last moment (at it is said in any apocalyptic book) and, the next morning, the millenium will be there.

2) The other way is the rational, responsible way, chosen by the Papacy and all learned, educated, sensible people of the time and of the following centuries. The goal is the same: to lead the world towards its end. But we know that the means are different, they are those which we have described in the preceding pages: reason, science, law, moderate politics (except as far as heresies are concerned, unfortunately). Choosing that way implies that you think that evil is everywhere, even in your own heart. God alone knows who the "saints" are. So everybody has to atone and to pay his or her debts. The war against the "wicked" is in vain. Everybody has to become better, to perform more good deeds, to use all his or her natural talents. The love of one's neighbour does not consist in killing the wicked, but in fulfilling the needs of one's neighbour, even the needs of the wicked.

This opposition between violent, magical eschatology, and rational, step-by-step eschatology helped structure political life in the West from the 11th to the 20th century. The Right and the Left (I mean the "pure" Right and the "pure" Left, which are both *revolutionary* movements, and which both wish *to move out of History*, either backwards or forwards, as shown by Karl Popper, instead of *improving* it) are the heirs of irrational, vio-

lent millenarism, whereas the democratic and liberal tradition is the heir to the spirit of the Papal Revolution. Marx, Lenin and Hitler are obviously representative of violent millenarism. The intellectual tradition which built the main concepts of democracy and liberalism are obviously representative of gradual, legal, scientific, rational millenarism, and of the spirit of the Papal Revolution.

9. A "spiritual", "prophetic" change

Let us now focus on the nature of this great civilizational change which created the "West". It seems that in this shift in eschatological perspective, these new moral duties of mankind, these new pastoral responsibilities of the Popes and of the whole Church, and finally these new principles of theology, none of these innovations are fully *explicable*, none of them could be *predicted*, none of them was *necessary*. We are obliged to acknowledge something such as a "prophetic" change, a "miracle". What happens here is a shift which is at the same time total and imperceptible. Ideas, values, institutions are completely re-organized over a few decades, but this is not done deliberately by anybody, and this can be understood only once it is done. How is such a phenomenon possible? Harold J. Berman himself explains the phenomenon in some convincing pages (Berman 1970, Introduction), but some further philosophical references can help to formulate more precisely what is at stake here: Henri Bergson's *The Two Sources of Morality and Religion*, Thomas Kuhn's *The Structure of Scientific Revolutions*, and Henri Atlan's *Entre le cristal et la fumée*. These authors tend to argue that all reality (whether natural or social) is perceived through certain *patterns* or *schemes* which alone give meaning and coherence to the dispersed, meaningless data of experience. Now the patterns themselves are not visible in general. "Normal science", Kuhn says, is systematic research made within the bounds of what he calls a "paradigm". When reality shifts, for whatever reason, more and more new data no longer fit into the pattern. Then the world begins to lose its meaning; it tends to become confused, opaque, incomprehensible. This impression of disorder may be made up, but the point is that, when this occurs, the process is *not* a "rational" one (in the sense of: conscious, logical) nor is it any kind of calculus or logical deduction. Rather, it comes by means of a *discontinuity*. Somebody *invents* a new pattern, a new paradigm through which the same confused data will become orderly.

Now then, this invention, just like artistic creation, is unexpected and not rationally understandable, which explains why many authors, such as Atlan, have paralleled such apparently different phenomena as *scientific discovery*, *artistic creation*, *entrepreneurial initiative*, as well as the *prophecy* of great religious or social reformers, or of great statesmen. In all these cases, they say, the same mental phenomenon is at stake, "invention" or "creation". In effect, the latter model is not *deduced* from the former, it was not contained or enclosed in it in any way. It is an "order made from chaos". The term "spiritual" designates the invisibility of this change, the internal causes of which one cannot see but the external effects of which one does see.

It is clear that the Papal Revolution was an event of this kind. Why did the Popes and their advisors think that Christ would no longer come back, that it was now up to men that He come, and, consequently, that they had the moral duty to transform society and to create the appropriate tools for such a purpose? Why did they see the social and political reality through such a scheme? Why were they tired of the old models? And why did this occur at that time, and not two or three centuries sooner or later? We are induced to speak of a "miracle". And this reminds us of some other such "miracles" in History.

10. The five miracles of Western History

They are not many, but the phenomenon recurs often enough in History to suggest some theoretical keys to the historical processes. For instance, I think that five "miracles" shaped the history of Western civilization:

1) The "Greek miracle" (a traditional expression), by which the City (*polis*) was invented, that is to say, a non-religious state, with the related principles of rule of law, equality in the eyes of the law, individual liberty under the law;

2) The "Roman miracle" (which is sometimes underrated), is the decisive improvement of the law by the invention of intellectual tools allowing the precise definition of private property, thus making possible the birth of the individual "ego", and hence, the rise of humanism;

3) The "Biblical miracle" is the invention of a new morality, with "love" or "mercy" extending beyond mere justice. If one loves, one can no longer admit evil (as the pagans did). One cannot be satisfied with merely repairing the wrong one has done, but also has the duty to root evil out of the world. One will remain guilty as long as some evil still exists in the world: that is the true meaning, I believe, of such a badly understood notion as "original sin". Such a moral revolution changes the sense of History, or in better terms, *creates* what we call "History": time becomes linear and no longer circular; it goes from a beginning, the Fall, towards an end, the defeat of evil. And this also changes the relation between spiritual power and temporal power, because the saints, not the state, have the responsibility of improving the world. The state is only an instrument. So the state, considered as a Babylon of sin, is desanctified; it can and should be strictly controlled. As shown by Graham Maddox (1997), this is the oldest root of democracy, which will emerge in the Middle Ages and extend into Protestant Modern times.

4) The Papal Revolution;

5) The Dutch, English, American democratic and liberal revolutions created the modern world. By relieving Christianity of most of its magic and superstitious aspects, and freeing individual thought, they created their own set of moral, political, social and economic values. To give a definition of the pattern or scheme they invented: they understood that *individual freedom* is not a source of *disorder*, but rather, the origin of *the most sophisticated orders* men can create, in democracy, in markets, in the critical methodologies of science. The fifth "miracle" of the West is the comprehension, for the first time in intellectual history, of the concept of "spontaneous" or "self-organizing" social order, or, to use Polanyi's words, of "the logic of liberty".

In each of these events, we can see the same *spiritual* element at work. There is a crisis in society; the world has become opaque and incomprehensible. But, at a certain moment, a new intellectual grid is introduced through which the world seems coherent and meaningful. From that moment on, almost all values, ideologies, institutions of the time

are reorganized and result in the creation of a new civilization.

I will give only two examples (I guess I could make the demonstration for any of the "miracles", but it would be much too long for this paper; see, for the first four "miracles", Nemo, 1998).

1. The circumstances in which the Roman jurists invented Roman law are remarkable. After the conquests, Rome had become the first multi-ethnic, multi-cultural state in history. The Roman magistrates had to deal with conflicts between citizens who had different customs and laws. They then made a long-range decision. In 242 B.C., they appointed a new magistrate, the *prætor peregrinus* ("praetor of foreigners"), specialized in trials opposing foreigners against one another or Romans against foreigners. This magistrate was allowed to address these trials by special laws, different from the old ethnic Roman law (the "Law of the Twelve Tables").

But, in the meantime, Rome had made a new conquest: Greece. So the Roman magistrates had come in contact with the Greek philosophers, especially the Stoics. The Stoics had moulded the concepts of "cosmopolis", based in the idea that, as a universal human nature exists, a universal natural law exists too. Positive laws are different in each city or ethnic group. But, "beneath" each ethnic law, one can find the same natural law. So, if you are a Roman magistrate trying to deal with a quarrel between a Gaul and a Syrian, you cannot appeal to the Gaul or the Syrian customs, for neither of them agree with, or even know, the customs of the adversary. But as they are men, they do have something in common, whether they know it or not. They have the same nature, implying a set of common rules to which they are ready to agree to if somebody conveys them. This is what the Roman magistrates did. Year after year (new prætors were elected every year), over roughly three centuries, they invented legal "*formulae*" (phrases) which were more universal because they were more and more *abstract*, independent from any particular ethnic rule. The important point is that Roman law was created, not simply because the Romans had conquered the world, but because there was an *idea* which orientated the process. But, on the other hand, no Stoic philosopher could have the *intention* of shaping Roman law, because none of them knew that a great and powerful multi-ethnic state would be created soon, achieving the first great figure of cosmopolis in History. So this great invention of law is not a truly purposeful design. Both human ideals and unexpected circumstances cooperate. In this sense, it is a "miracle".

2. Modern scholars, from the 17th century onwards, and especially during the latter decades, have illuminated many historical steps in the construction of the Bible. At every step, some contingent circumstance seemed to prevail. It is because the Persians had vanquished the Babylonians that Judaea became a Persian *satrapy* (province). It is because the Persian governors, in Judaea as well as in the whole Persian Empire, needed to know how to administer the Jews, that they needed a written code of Judaic law, and this is why they took the initiative of writing the Torah. The final text of the Torah was drawn up under the auspices of two Judeo-Persian governors, Ezra and Nehemiah, obviously with a political purpose. We find the same kinds of contingent, sometimes anecdotal circumstances at almost every other step (the separation between Northern and Southern kingdoms, due to the Assyrian conquest, causing the elaboration of two different versions, "elohist" and "yahvist", of the sacred history; later, the fall of the North causes the incomplete, illogical merging of the two texts and King Josias' reformation giving rise to a new

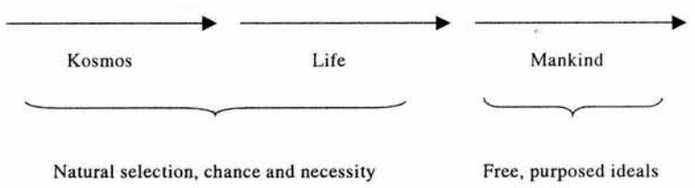
text, Deuteronomy, etc.). So the rise of one of the most sacred texts of mankind must be given secular and prosaic, rather than religious, explanations. One can no longer believe that the Torah was written by God's finger breaking up the clouds, nor by Moses directly hearing God's voice in the burning bush. Scholars have elaborated many such sharp, destructive new explanations. The Gospels, too, can no longer be thought of as direct testimonies; these late texts, completed decades after Christ's death, reflect the slowly grown and liturgically organized creeds of some defined communities. Certain other testimonies, defended by other groups, have been definitively lost or underrated. The closing of the canons both of the Old and New Testament bear an impression of fights between rival groups, and sometimes it is visible that the strongest, and not the wisest, has won.

Nevertheless, to suggest that the Bible is only a "by-product" of a series of such contingent events would be entirely wrong. As Emmanuel Levinas once told me, the true miracle is not that God showed His face to a small number of men, while hiding it to the others. On the contrary, the really striking miracle is that, given that the text is the product of a process which took up centuries, in which hundreds of independent, variously motivated people interfered, *most of these various additions nevertheless converged.* The miracle is that the final result of this non-deliberate process was finally so coherent in depth, as the work of commentators along centuries has continuously shown. Once more, this paradoxical convergence of dispersed intellectual and practical initiatives can be explained only if we acknowledge its "spiritual" nature, namely, the role of the Jewish and Christian *prophets,* who did nothing but conveying the new morals and the new vision of the future of mankind in few but explicit words. Since these words were set forth, an *idea* was present in the world which gave all the other biblical writers the same internal patterns through which they could "see the situation" in a similar way, so that almost any of the further contingent historical events would be seen to fit into the same drama rather than being foreign to it.

11. Reason and irrationality in History

This suggests a final reflection about History. Today, we know that the whole world is evolving, not only human societies. Biology since Darwin, astronomy in more recent years, have proved that life and physical *kosmos* have their own "history". But, in any other field than human life, evolution is the mere result of "chance and necessity". One can see in it no intentions whatsoever. Surely, in many regards, this remains true even within human history itself, where many things seem to be shaped only by evolutionary, trial-and-error processes, according to the logic demonstrated by Hayek. But I think that this is not sufficient to explain history in depth (and this could be the limit of Hayek's philosophy). One can see that sometimes, at certain privileged moments, such as those which I dealt with previously, history is *guided, oriented* by human initiatives. From time to time, some illuminated men can "see the future", and, only by "seeing" it, *create* it. As soon as man appears on the scene, it seems that he can cooperate with "chance and necessity"or with God in creating the world. Creation is no longer a solitary, anonymous device, now man takes part in it.

If we accept this idea, progress in History would not only be a random process. It



could be, to a certain degree, an intelligent, purposeful process, guided by a deliberate reflection on the past, present and future of mankind. But our description of the five "miracles" of Western history has shown us that this creation, whatever it could be, is not a *logical* process. For the new vision which cleaves history is not *deduced* from the old visions. It arrives owing to a strange "alchemy" which no Cartesian intelligence can explain, because though this alchemy changes the vision, it is not itself visible. As St John of the Cross demonstrated (possibly better than Hegel), "spirit" goes forward only through "night", so, in a sense, "spirit" *is* "night". As progress in History – whether it be science, art, or moral, political and social reforms – is a *discontinuous* process, we should think of it as a *nocturnal* process. This condemns any positivist vision of history and of the history of sciences, in the sense of Auguste Comte or of the Vienna circle.

This changes, finally, our vision of *rationality* itself. If we acknowledge that history advances by intellectual leaps, and if we acknowledge in particular that the rational civilization *par excellence*, Western civilization, was not the fruit of any clear-and-distinct reasoning, but was born from a new vision of the world prophetically proposed by the Papal Revolution, we must necessarily elaborate a new, wider concept of reason. In the narrow sense of this notion, the rational is only that which "fits into" a given scheme or pattern. That which falls outside the framework is irrational and seems to be either foolish or wicked. But we know that it is ignorant to think that somebody who disagrees with you is necessarily foolish or wicked. Erasmus wrote a "Praise of Folly" precisely to show that, very often, a new truth comes through a "foolish" statement. Kuhn pointed out the same: scientific progresses are achieved by "revolutions" which can appear but foolish to the "normal" scientists; but, if the new paradigm proves to be better than the old one, eventually everybody will think that it was the opponents of the revolution who were foolish. The important point is that, at the time when a new paradigm is proposed, no agreement is conceivable among scientists. We can say that the logic of scientific discovery creates a gulf between minds: if one of them is rational, the other one must be irrational, and vice versa. But every philosopher will agree that science is rational on the whole. Therefore it seems that we must *integrate the gap within our very concept of reason*. Not only what fits into a given scheme is rational, but the process of inventing new schemes belongs to reason in a wider sense.

The question is: how can the final truth of the process be determined? For sure, some or many minds are *really* foolish. Every thesis or work which seems foolish will not necessarily, in retrospect, prove to have been a positive step in the intellectual, scientific, artistic, moral progress of mankind. So we need a criterion by which to judge, and I would

argue that the only valuable criterion is history itself. Only through time and empirical events do things become clear. As Popper demonstrated, the ultimate proof of a scientific "revolutionary" theory is the fact that nobody succeeds in proving it wrong. But nobody can know in advance whether it will be refuted or rebutted or not. Similarly, the true value of a social vision is its social fruitfulness, which comes to light only through time. What makes the difference between Peter Abelard, founder of the scolastic method, and his adversary St Bernard of Claivaux, supporter of the old symbolic and poetic exegesis of the Bible? Nothing could during their lifetime. Both were exceptionally intelligent, brilliant, persuasive, while each of them was a wicked man and a fool in the eyes of the other. Only now, after centuries, do we know that St Bernard was wrong to oppose Abelard so fiercely, because we have witnessed the fruitfulness of Western scientific civilisation. The controversy can be settled now, but *analytical* reason was then of no avail. Truth cannot be *constructed*, it has to be *found*.

This means that we cannot spare *time* and *night*. In this sense, history *is* reason, as well as analytical thought. We must aknowledge the limits of our analytical reason in order to enhance the powers of our wider reason; "Esprit", "Geist", "Spirit" designate this nocturnal part of Reason. The paradox is that the spirit creates the world, while our analytical, positive reason does not comprehend the spirit, nor can substitute for it. That is the true *rational* base of freedom.

References

Atlan, Henri: *Entre le cristal et la fumée*, Paris, Editions du Seuil, 1979.

Berman, Harold J.: *Law and Revolution. The Formation of the Western Legal Tradition*, Harvard University Press, 1983.

Bergson, Henri (1935): *The Two Sources of Morality and Religion (Les Deux Sources de la morale et de la religion)*, University of Notre Dame Press, 1977

— (1946): "Introduction to metaphysics", in *The Creative Mind (La pensée et le mouvant)*, New York, Greenwood Press, 1968

Kuhn, Thomas S.: *The Structure of Scientific Revolutions*, University of Chicago Press, 1996.

Maddox, Graham: *Religion and the Rise of Democracy*, Routledge, London & New York, 1996.

Nemo, Philippe (1996): *Athènes, Rome, Jérusalem: trois sources de la civilisation occidentale*, in "L'Union européenne et les Etats-nations. The European Union and the Nation States", ESCP Press.

— (1998): *Histoire des idées politiques dans l'Antiquité et au Moyen Âge*, Presses Universitaires de France.

Popper, Karl Raimund: *Open Society and its Enemies*, 1945, 5th revised edition, Princeton University Press, 1971.

The Picture Theory of Reason*

J. C. Nyíri

The picture which philosophy today conjures up when addressing the nature of rational thinking is that of a verbal process, spoken, written, or silent. In contrast to this, I would like to show that rational – coherent, logical – thought and communication essentially involve non-verbal symbols too. Among these, *visual* symbols are the most important. By visual symbols I mean mental images as well as public pictures, diagrams, and models, and what I will particularly aim at demonstrating is that the growing abundance and increasingly easy production of pictures on the screen radically improve our capacity to develop theories of visual imagery as well as to develop an understanding of their central role in our cognitive economy.

§ 1. From Plato to Wittgenstein

Philosophers, just like everyone else, have at all times enjoyed more or less realistic impressions of what thinking feels like: they have heard themselves thinking in words, but also they have experienced themselves as thinking via mental images. But when they attempted to express, formulate, and communicate their impressions, they had no choice but to use a one-sided medium: the medium of language. And since, before the days of computer graphics, philosophers seldom actually dealt with pictures, had little practical knowledge of them, and had no terminology to talk about them, they ended up either by conceiving of mental images in terms of verbal language, or by suppressing the notion of thinking in images entirely. In particular, they repudiated the notion that images might play a role in abstract reasoning. Plato and Aristotle were certainly haunted by the belief that thinking is a visual activity; but this belief became in later times no more than an underground current in philosophy, surfacing, temporarily, with the British Empiricists, and resurgent, again, today.

Plato chose the words ***idea*** and ***eidos*** to designate the abstract objects of thought. These words, which he used alternately, mean "form" or "shape". Both ***idea*** and ***eidos*** come from the verb ***idein***, "to see"; from ***eidos*** there descends the word ***eidolon***, "the visible image". Though ***eidos*** is not etymologically related to ***eikon*** – "likeness", "picture" – the acoustic and semantic proximity between the two does suggest a kind of relatedness, and Plato is not always willing, or able, to avoid that suggestion. Thus in the *Euthyphro*,

* I am deeply indebted to Csaba Pléh, Professor of Psychology at the Universities of Budapest and Szeged, for innumerable discussions relating to the topics of this paper, and his unrelenting efforts to keep me alert to new developments in cognitive psychology. I am grateful to Barbara Tversky for valuable comments she made on the text of my paper as presented in Kirchberg.

B. Brogaard, B. Smith (Hg.), *Rationality and Irrationality / Rationalität und Irrationalität*, S. 242–266.

where Socrates wishes to learn what the idea of holiness is, so that he "may keep [his] eye fixed upon it and employ it as a model (***paradeigma***)",[1] or in the *Meno*, according to which the soul, prior to being born, has actually *seen* the ideas,[2] or in the *Phaedrus*, with its fable of "glorious and blessed sights in the interior of heaven" and "a vision of the world beyond".[3] But of course at the very same place in the *Phaedrus* Plato tells us that "essences", i.e. ideas, are "formless, colourless, intangible, perceived by the mind only". And in the *Republic* we learn that "ideas can be thought but not seen".[4] Platonic ideas are not images; they are abstract word-meanings. Plato's philosophy emerges under the impact of the rise of *alphabetic literacy*.[5] Pre-literal narrative language, as Herder already stressed,[6] is inherently metaphoric, feeding on, and fostering, images; with the rise of

1. 6e – translation by Harold North Fowler.
2. 81c. – On Plato's conflicting views on conveying knowledge by graphic means see Petra Gehring – Thomas Keutner – Jörg F. Maas – Wolfgang Maria Ueding, eds., *Diagrammatik und Philosophie*, Amsterdam: Rodopi, 1992, pp. 7 and 15ff.
3. 247a–c – translated by B. Jowett.
4. 507c – translated by Paul Shorey.
5. In a telling passage of the *Philebus* Plato compares the soul to a *book*, adding however that besides the "scribe" who writes within us there is also "another artist, who is busy at the same time in the chambers of the soul": "The painter, who, after the scribe has done his work, draws images in the soul of the things which he has described" (39a–b, Jowett transl.).
6. Let me refer to his early piece "Über die neuere Deutsche Litteratur. Erste Sammlung von Fragmenten" (1766 – I am quoting from Herders *Sämmtliche Werke*, ed. Bernhard Suphan, vol.1, Berlin: Weidmannsche Buchhandlung, 1877). In the childhood phase of language "[sprach] man noch nicht ..., sondern tönete; ... man [dachte] noch wenig ..., aber [fühlte] desto mehr ...; und also nichts weniger als schrieb" – "Man sang also, wie viele Völker es noch thun und wie es die alten Geschichtschreiber durchgehends von ihren Vorfahren behaupten" – "Das Kind erhob sich zum Jünglinge ... die Lebens- und Denkart legte ihr rauschendes Feuer ab: der Gesang der Sprache floß lieblich von der Zunge herunter, wie dem Nestor des Homers, und säuselte in die Ohren. Man nahm Begriffe, die nicht sinnlich waren, in die Sprache; man nannte sie aber, wie von selbst zu vermuthen ist, mit bekannten sinnlichen Namen; daher müssen die ersten Sprachen Bildervoll, und reich an Metaphern gewesen seyn" (p. 153). And on the next pages: "je mehr bürgerliche und abstrakte Wörter eingeführet werden, je mehr Regeln eine Sprache erhält: desto vollkommener wird sie zwar, aber desto mehr verliert die wahre Poesie" (p. 154). – "Das hohe Alter [der Sprache] weiß statt Schönheit blos von Richtigkeit. ... Je mehr die Grammatici den Inversionen Fesseln anlegen; je mehr der Weltweise die Synonymen zu unterscheiden, oder wegzuwerfen sucht, je mehr er statt der uneigentlichen eigentliche Worte einführen kann; je mehr verlieret die Sprache Reize: aber auch desto weniger wird sie sündigen. ... Dies ist das /dies wäre ein Philosophisches/ Zeitalter der Sprache" (p. 155). That is, according to Herder the words of oral language have figurative, metaphorical, non-literal meanings; written language is abstract, philosophic, non-figurative. – Let me also cite from the lectures Nietzsche gave on Greek literature in 1872, at the University of Basel (my source being: *Nietzsche's Werke*, Leipzig: Alfred Kröner. Vol.XVIII. *Philologica.* Second volume. *Unveröffentlichtes zur Litteraturgeschichte, Rhetorik und Rhythmik*, 1912). Nietzsche refers to "die Tropen, die uneigentlichen Bezeichnungen", and continues: "Alle Wörter aber sind an sich und von Anfang an, in Bezug auf ihre Bedeutung, Tropen" (p. 249). Nietzsche's explanation of synecdoche, metaphor, and metonym: "die Tropen treten nicht dann und wann an die Wörter heran, sondern sind deren eigenste Natur. Von einer 'eigentlichen Bedeutung', die nur in speciellen Fällen übertragen würde, kann gar nicht die Rede sein" (p. 250). In these lectures Nietzsche strives to demonstrate the essentially *oral* character of Greek literature. His thesis, then: the

written language, however, not only spoken language, but also the language of images became relegated to an inferior position.

The history of Western philosophy is a history of recurrent clashes between the experience of imagery on the one hand, and the experience of written language on the other; with written language invariably being the victor. Aristotle's thesis, according to which "the soul never thinks without an image" – ***phantasma*** – must have remained barren in the context of *De anima*, dominated as it was by the metaphor of the mind as a *writing-table* (***grammateion***).[7] Locke's famous difficulty, described in Book 4, chapter 7 of his *Essay* – namely that it does indeed "require some pains and skill to form the general idea of a triangle, (which is yet none of the most abstract, comprehensive, and difficult,) for it must be neither oblique nor rectangle, neither equilateral, equicrural, nor scalenon; but all and none of these at once. In effect, it is something imperfect, that cannot exist; an idea wherein some parts of several different and inconsistent ideas are put together" – definitely expresses some deep ambiguity. Ideas seem to be of a pictorial nature (otherwise the general idea of a triangle would not cause embarrassment) but also they must permit of non-pictorial dimensions (since as *generic* pictures, Locke implies, they *cannot* exist). In fact in the *Essay* there is a marked tendency to equate ideas with single *written words*. The mind, at birth, is like a "white paper, void of all characters, without any ideas"; when describing the doctrine of *stamped*, or *imprinted*, innate characters,[8] it is only the innateness Locke takes issue with. Berkeley, insisting that ideas are indeed images, still believes that generic mental images are inconceivable.

In the course of my talk I will cite arguments in favour of the existence of generic images, and come back to Berkeley's criticism of Locke. Of the British Empiricists, it is Hume whose views of the thinking process are most unequivocally imagistic. "The term 'idea'", as Russell's *The Wisdom of the West* puts it, "is here to be understood in the literal Greek sense of the word. Thinking, for Hume, is picture thinking, or imagining, to use a Latin word which originally meant the same."[9] Hume's observations are still today useful when it comes to explaining how pictures can mean what they mean, and I will return to those observations later. In the event, however, Hume too remained a prisoner of Gutenberg, just as did every modern philosopher before Nietzsche. Let me mention here the *Critique of Pure Reason*, the grand task of which was to provide a synthesis of sensi-

words of pre-literal language are, without exception, tropes – words expressing images. And a last quote from this material, here Nietzsche quoting: "richtig Jean Paul, Vorschule der Aesthetik: 'Wie im Schreiben Bilderschrift früher war als Buchstabenschrift, so war im Sprechen die Metapher, insofern sie Verhältnisse und nicht Gegenstände bezeichnet, das *frühere* Wort, welches sich erst allmählich zum *eigentlichen Ausdrucke* entfärben musste. Das Beseelen und Beleiben fiel noch in Eins zusammen, weil noch Ich und Welt verschmolz. Daher ist jede Sprache in Rücksicht geistiger Beziehungen ein Wörterbuch erblasseter Metaphern'" (pp. 264f.).

7. Aristotle, *On the Soul*, 431a and 430a, transl. by J. A. Smith. *The Complete Works of Aristotle: The Revised Oxford Translation*, ed. by Jonathan Barnes, Princeton: Princeton University Press, 1984. I am indebted to István Bodnár for his generous help over issues regarding Plato and Aristotle.
8. John Locke, *An Essay Concerning Human Understanding*, Book 2, chapter 1, section 2; and Book 1, chapter 1, sections 1 and 5.
9. Bertrand Russell, *The Wisdom of the West*, London: Macdonald, 1959, p. 225.

bility and conceptual knowledge, a task the author, constrained by the limitations of a linear text unsuited to deal with the facts of pictorial thinking, could clearly not realize. Let me just refer to the Schematism of the Pure Conceptions of the Understanding – "th[e] representation of a general procedure of the imagination to present its image to a conception, I call the schema of this conception"[10] –, or indeed to the fundamental Kantian definitions of the understanding itself, like for instance: *Verstand ist das Vermögen, den Gegenstand sinnlicher Anschauung zu denken*[11]. The first philosopher who in fact used pictures – drawings – to illustrate some of the points he wanted to make about seeing and imagining, was the later Wittgenstein,[12] a philosopher for whom written language has lost its spell,[13] and whose views on mental imagery were rather less unequivocal then is generally supposed. I will return to Wittgenstein shortly.

Russell's *The Wisdom of the West* was published in 1959. Wittgenstein got short shrift in it, his fondness for using pictures not being mentioned at all – even though this book was an attempt, as it was put in the foreword, "wherever this seemed feasible, to translate philosophic ideas, normally expressed only in words, into diagrams that convey the same information by way of geometrical metaphor".[14] The attempt failed dismally. Clearly, Russell, and his editor, themselves suffered from the disability which he had diagnosed in his fellow philosophers forty years earlier, in "On Propositions", where he wrote: "The habit of abstract pursuits makes learned men much inferior to the average in the power of visualization, and much more exclusively occupied with words in their 'thinking'."[15]

10. A 140 – translation by J. M. D. Meiklejohn.
11. A 51.
12. As Andreas Roser writes: "Der frühe philosophische Gebrauch von Bildern ließe sich vielleicht am ehesten noch durch wortsprachliche Beschreibungen von allegorischen oder gleichnishaft verwendeten Bildern darstellen – etwa in einigen Gleichnissen Platons. Doch eben diese – obwohl sie illustrierbar gewesen wären – wurden nicht illustriert. Und Philosophen in der Tradition Platons wären günstigstenfalls geneigt, den Versuch eines solchen Illustrationsvorhabens zu belächeln. Wird man auch in diesen frühen Epochen der Philosophie kaum fündig werden, so ist es um die *philosophische Verwendung* von Bildern im Mittelalter bereits besser bestellt. – In einigen Schriften des Nikolaus v. Kues oder später – in der Renaissance – in den änigmatischen Bildern des Giordano Bruno oder in den naturphilosophischen Schriften Descartes' finden sich Bilder in sprachphilosophischer Funktion, um mit ihrer Hilfe etwas darzustellen, das sich mit Hilfe der Wortsprache nur unangemessen oder unvollständig hätte darstellen lassen. Doch dies sind die seltenen Ausnahmen. – Die überwiegende Anzahl aller philosophischen Arbeiten verzichtet – zumeist ohne Angabe von Gründen – auf jeden Gebrauch von Zeichnungen oder Grafiken zur Klärung *philosophischer* Probleme." (Roser, "Gibt es autonome Bilder? Bemerkungen zum grafischen Werk Otto Neuraths und Ludwig Wittgensteins", *Grazer Philosophische Studien* 1996/97, pp. 13f.)
13. See my "Wittgenstein as a Philosopher of Secondary Orality", *Grazer Philosophische Studien* 52 (1996/97), pp. 45–57.
14. *Op. cit.*, p. 5. Russell did not actually write the *Wisdom of the West*. As he puts it in his *Autobiography*, the book was "taken" (by the "editor" Paul Foulkes) from his *History of Western Philosophy*, "but ironed out and tamed". Still, as Russell stresses, he *did* like the illustrations in it. (*The Autobiography of Bertrand Russell*, London: George Allen and Unwin, vol. 2, 1968, pp. 223f. – For the real story of the book, see Carl Spadoni, "Who Wrote Bertrand Russell's *Wisdom of the West*?", *Papers of the Bibliographical Society of America*, 80:3 [1986].)
15. Bertrand Russell, "On Propositions: What They Are and How They Mean" (1919). *Aristotelian*

Russell's view will be echoed by H.H. Price in 1953, in his *Thinking and Experience*. As Price writes: "We have the misfortune to live in the most word-ridden civilization in history, where thousands and tens of thousands spend their entire working lives in nothing but the manipulation of words. The whole of our higher education is directed to the encouragement of verbal thinking and the discouragement of image thinking. Let us hope that our successors will be wiser, and will encourage both."[16]

§ 2. Visual Thinking

By way of encouraging thinking with images, let me recall here a puzzle readers of Arthur Koestler will be familiar with.[17]

> One morning, exactly at sunrise, a Buddhist monk began to climb a tall mountain. The narrow path, no more than a foot or two wide, spiralled around the mountain to a glittering temple at the summit. – The monk ascended the path at varying rates of speed, stopping many times along the way to rest and to eat the dried fruit he carried with him. He reached the temple shortly before sunset. After several days of fasting and meditation, he began his journey back along the same path, starting at sunrise and again walking at variable speeds with many pauses along the way. His average speed descending was, of course, greater than his average climbing speed. – Prove that there is a single spot along the path the monk will occupy on both trips at precisely the same time of day.

Koestler liked to put this problem to his friends. The conclusion the mathematically minded among them tended to come to was that it would be an unlikely coincidence if the monk happened to be at the same spot at the same time of the day in the course of such two

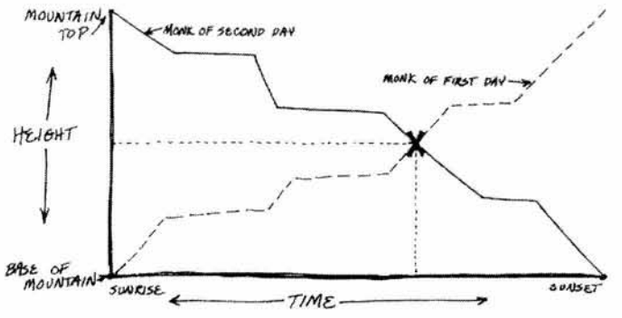
Society Supplementary Volume, 2, pp. 1–43. I am here quoting from J.G. Slater (ed.), *The collected papers of Bertrand Russell*, Volume 8: *The Philosophy of Logical Atomism and Other Essays, 1914-19*. London: George Allen & Unwin, 1986, pp. 284f. The passage in full: "If you try to persuade an ordinary uneducated person that she cannot call up a visual picture of a friend sitting in a chair, but can only use words describing what such an occurrence would be like, she will conclude that you are mad. (This statement is based upon experiment.) I see no reason whatever to reject the conclusion originally suggested by Galton's investigations, namely, that the habit of abstract pursuits makes learned men much inferior to the average in the power of visualizing, and much more exclusively occupied with words in their 'thinking'". When Professor Watson says: 'I should throw out imagery altogether and attempt to show that practically all natural thought goes on in terms of sensori-motor processes in the larynx (but not in terms of imageless thought)' (*Psychological Review*, 1913, p. 174n.), he is, it seems to me, mistaking a personal peculiarity for a universal human characteristic." Later in the paper Russell writes: "The 'meaning' of images is the simplest kind of meaning, because images resemble what they mean, whereas words, as a rule, do not", p. 292.

16. London: Hutchinson's Universal Library, p. 252. Some pages later Price makes the telling remark: "some people are almost incapable of drawing", p. 258.
17. See Koestler, *The Act of Creation*, London: Hutchinson, 1964, pp. 183f. Koestler refers to the June 1961 issue of *Scientific American* as his source, but remarks that the problem originates with the psychologist Carl Duncker.

different journeys. Others however *saw* the solution. These are the words of "a young woman without any scientific training":

> I tried this and that, until I got fed up with the whole thing, but the image of the monk in his saffron robe walking up the hill kept persisting in my mind. Then a moment came when, superimposed on this image, I saw another, more transparent one, of the monk walking down the hill and I realized in a flash that the two figures must meet at some point some time – regardless at what speed they walk and how often each of them stops. Then I reasoned out what I already knew: whether the monk descends two or three days later comes to the same; so I was quite justified in letting him descend on the same day, in duplicate so to speak.[18]

Even for the graphically minded it is not entirely trivial to visualize this solution. Robert McKim for instance, in his *Experiences in Visual Thinking*, manages to add a rather con-



fusing diagram to an otherwise useful text.[19] The ideal expedient suggesting itself here, with the means at our disposal these days, seems to be some appropriate *animation*.

Mental imagery itself appears to be a matter of dynamic, rather than static, pictorial representations. And the rudimentary capacity of thinking *through* images, of thinking *directly* with images, without verbal mediation, seems to belong to our *biological* makeup. Merlin Donald, in his *Origins of the Modern Mind*, distinguishes three evolutionary transitions in the development of humankind. The first transition, from apes to *Homo erectus*, was characterized by "the emergence of the most basic level of human representation, the ability to mime, or re-enact, events". The second transition, from *Homo erectus* to *Homo sapiens*, completed the biological evolution of modern humans. "The key event during this transition", writes Donald, "was the emergence of the human speech system, includ-

18. *Ibid.*, p. 184.
19. Robert H. McKim, *Experiences in Visual Thinking* (1972), Boston: PWS Publishing Company, 1980, p. 3. The diagram gets slightly worse in McKim's other volume, *Thinking Visually: A Strategy Manual for Problem Solving* (Palo Alto, CA: Dale Seymour Publications, 1980, p. 4) – basically the same book in a different guise.

ing a completely new cognitive capacity for constructing and decoding narrative."[20] As Donald emphasizes: "Speech provided humans with a rapid, efficient means of constructing and transmitting verbal symbols; but what good would such an ability have done if there was not even the most rudimentary form of representation already in place? There had to be some sort of semantic foundation for speech to have proven useful, and mimetic culture would have provided it."[21] The third transition was "recent and largely nonbiological, but in purely cognitive terms it nevertheless led to a new stage of evolution, marked by the emergence of visual symbolism and external memory as major factors in cognitive architecture. External symbolic storage", Donald stresses, "must be regarded as a *hardware* change in human cognitive structure, albeit a nonbiological hardware change."[22] To this last transition Donald allots "three broadly different modes of visual symbolic invention", which he designates as "pictorial, ideographic, and phonological".[23] Of these, the pictorial mode emerged first; and the point Donald makes is that this signaled the beginnings of "a new cognitive structure",[24] already enabling some primitive forms of "analytic thought", i.e. "formal arguments, systematic taxonomies, induction, deduction".[25]

Now if the role pictures play in our cognitive activities is so considerable, why did philosophers, up to the twentieth century, practically never make use of them? The answer is that, almost throughout recorded history, the production and duplication of pictures was a much more cumbersome and unreliable undertaking than the writing down, and copying, of texts. In pre-literate times pictures obviously fulfilled an indispensable function in the storage and communication of collective knowledge. With the emergence of phonetic writing, pictures receded into the background, although they did still function as mnemonic devices. The Romans used simple pictures, called *emblems*, to help them overcome the inherent visual deficiency of their scripts.[26] They recalled specific parts of a text by remembering the particular emblem placed against it in the margin. In early medieval manuscripts illustrations helped readers to find the part of the text they were looking for. Applied in this manner, pictures had a merely auxiliary function; and even that was lost with the introduction of word separation. This invention gave written Latin an ideographic value without sacrificing the inherent advantages of a phonetic alphabet.[27] Pictures now ceased to be needed as visual aids. And prior to printing they could not become aids to the communication of knowledge. Since they were inevitably distorted in the copying process, information could not be preserved by them. There are some enlightening passages by Pliny the Elder in his *Natural History*, written in the first century of our era, describing what can only be regarded as the ultimate failure of Greek botany as a science.

20. Merlin Donald, *Origins of the Modern Mind: Three Stages in the Evolution of Culture and Cognition*, Cambridge, Mass.: Harvard University Press,1991, p. 16.
21. *Ibid.*, p. 199.
22. *Ibid.*, p. 17.
23. *Ibid.*, p. 278.
24. *Ibid.*, p. 284.
25. *Ibid.*, p. 273.
26. Paul Saenger, "Silent Reading: Its Impact on Late Medieval Script and Society", *Viator* 13 (1982), p. 372.
27. *Ibid.*, p. 377.

ence. Let me here quote the dramatic summary of those passages given by Ivins:

> The Greek botanists realized the necessity of visual statements to give their verbal statements intelligibility. They tried to use pictures for the purpose, but their only ways of making pictures were such that they were utterly unable to repeat their visual statements wholly and exactly. The result was such a distortion at the hands of the successive copyists that the copies became not a help but an obstacle to the clarification and the making precise of their verbal descriptions. And so the Greek botanists gave up trying to use illustrations in their treatises and tried to get along as best they could with words. But, with words alone, they were unable to describe their plants in such a way that they could be recognized – for the same things bore different names in different places and the same names meant different things in different places. So, finally, the Greek botanists gave up even trying to describe their plants in words, and contented themselves by giving all the names they knew for each plant and then told what human ailments it was good for. In other words, there was a complete breakdown of scientific description and analysis once it was confined to words without demonstrative pictures.[28]

Picture printing was invented around 1400 A.D. Ivins argues that this was a much more revolutionary invention in the history of communication than that of typography half a century later. Pictures became more or less exactly repeatable. However, they were still a long way from being faithful copies of particular natural objects; indeed the very demand for faithful representations emerged only gradually in the course of the fifteenth century. The so-called *Pseudo-Apuleius*, a printed version of a ninth-century botanical manuscript, published just after 1480 at Rome, contains woodcuts that are careless copies of the manuscript illustrations, and could of course not be of any practical use. Already just a few years later the German herbal *Gart der Gesundheit* is printing woodcuts based on expert drawings of the original plants. However, neither woodcuts, nor etchings or engravings, could aim at complete faithfulness. Ivins points

28. William Ivins, Jr., *Prints and Visual Communication*, London: Routledge and Kegan Paul, 1953, p. 15.

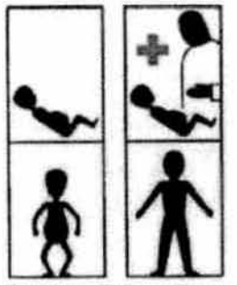
out that when Lessing wrote his famous treatise on the Laocoon group, he did not, because he could not, have reliable illustrations at his disposal. "Each engraver", writes Ivins, "phrased such information as he conveyed about [the statues] in terms of the net of rationality of his style of engraving. There is such a disparity between the visual statements they made that only by an effort of historical imagination is it possible to realize that all the so dissimilar pictures were supposed to tell the truth about the one identical thing. At best there is a family resemblance between them."[29] Until the age of photography, as Ivins stresses, there existed no technology of exactly repeatable pictorial representations of particular objects.

The head of Laocoon. Engraving around 1527, woodcut 1544, etching 1606. After Ivins.

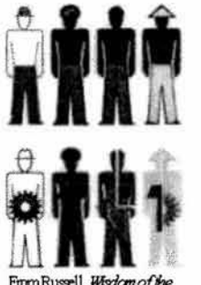
Let us add that even *after* the advent of photography and until the emergence of computer graphics, texts could be manipulated with much greater ease, both by the author and especially by the printer, than could pictures. Thus scientists, and certainly scholars, often had to make do with texts even when graphical representations would have been called for. Sometimes this was felt to be a possible loss. Already Bacon wrote:

> Emblem reduceth conceits intellectual to images sensible, which strike the memory more ... Aristotle saith well, 'Words are the images of cogitations, and letters are the images of words.' But yet it is not of necessity that cogitations be expressed by the medium of words. For whatsoever is capable of sufficient differences, and those perceptible by the sense, is in nature competent to express cogitations.

Although, as he puts it, "words and writings by letters do far excel all other ways", Bacon still finds it necessary to investigate the possible uses of "characters real".[30] However, such investigations could not really begin before the late twentieth century. Even in the 1920s and 30s they turned out to be technologically premature, as is revealed by the failed experiments of Otto Neurath. Neurath was working towards an "International System Of Typographic Picture Education", abbreviated as *isotype*, an interdependent and interconnected system of images, to be used together with *word languages*, yet having a visual logic of its own. Isotype would be two-dimensional,[31] using distinctive conventions, shapes, colours, and so on. "Frequently it is very hard", Neurath wrote, "to say in words what is clear straight away to the eye. It is unnecessary to say in words what we are able to make clear by pictures".[32] Neurath particularly stressed that the elaboration of his picture language was meant to serve a broader task, that of establishing an international encyclopaedia of common, united knowledge – the "work of our time", he said.[33] However, he never even came near to realizing such lofty aims. The icons elaborated within the framework of the isotype program have served as models for those international picture signs we today daily encounter at airports and railway stations, but – because they are so crude, and so cumbersome to produce – they could not form the basis of a true visual *language*. Even so, Neurath's program does raise some interesting *theoretical* questions with respect to pictures. Central among these is the *natural resemblance* vs. *conventionality* issue – an issue at the heart of Wittgenstein's philosophy of pictures.



Go to the medical man, if your baby has Rachitis
From Neurath, *International picture language* (1936)



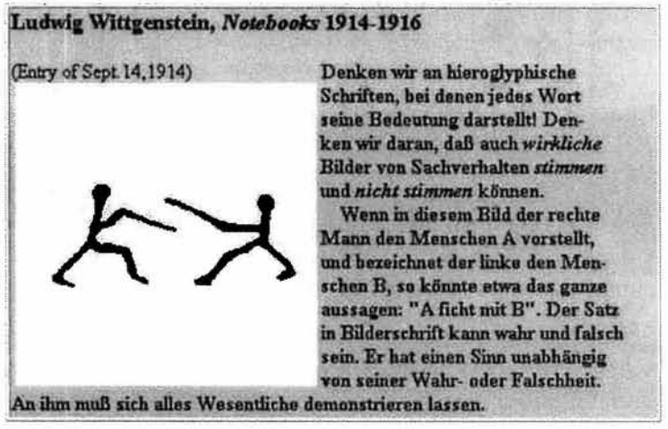
From Russell, *Wisdom of the West*. "A sample of O. Neurath's use of pictorial symbols to overcome problems of communication"

§ 3. From Wittgenstein to Goodman

Wittgenstein's philosophy of pictures is commonly regarded as comprising two contrasting positions. The *Tractatus* is taken to argue for *a picture theory of meaning*,[34] summed up by Wittgenstein's dictum: "The proposition is a picture of reality ... In order to understand the essence of the proposition, consider hieroglyphic writing, which pictures the

29. *Ibid.*, p. 89.
30. Francis Bacon, *The Advancement of Learning*, Oxford: Clarendon Press, 1974, pp. 130f.
31. "The writing or talking language is only of 'one expansion' – the sounds come one after the other in time, the word-signs come one after the other on paper, as for example the telegram signs on a long, narrow band of paper. The same is true in books – one word over another in the line under it has no effect on the sense. But there are languages of 'two expansions'." Otto Neurath, *International Picture Language*, London: 1936, repr. University of Reading: Department of Typography & Graphic Communication, 1980, p. 60.
32. *Ibid.*, p. 26.
33. *Ibid.*, pp. 65 and 111.
34. A subtle and imaginative analysis of Wittgenstein's idea of a directly depicting language is given in Barry Smith, "Characteristica Universalis", in: K. Mulligan, ed., *Language, Truth and Ontology*, Dordrecht: Kluwer, 1992.

facts it describes. And from it came the alphabet without the essence of the representation being lost. This we see from the fact that we understand the sense of the propositional sign, without having had it explained to us." The later Wittgenstein is interpreted as holding a *use theory of pictures*, according to which pictures by themselves do not carry any meaning; they acquire meaning by being put to specific uses and by being applied in specific contexts. Those uses are embedded in, and those contexts are determined by, *language*; pictures are subservient to words, and indeed not even *mental images* mean by virtue of resemblance. "'The image must be more like its object than any picture', Wittgenstein has his imaginary antagonist say in the *Philosophical Investigations*. 'For, however like I make the picture to what it is supposed to represent, it can always be the picture of something else as well. But it is essential to the image that it is the image *of this* and of nothing else.' Thus one might come", adds Wittgenstein, "to regard the image as a super-likeness", *die Vorstellung als Über-Bildnis*.[35] The *locus classicus* of the later Wittgenstein's arguments against pictures as natural signs is of course §139 of *Philosophical Investigations*. I am quoting the crucial lines:



> What really comes before our mind when we *understand* a word? – Isn't it something like a picture? Can't it *be* a picture? – Well, suppose that a picture comes before your mind when you hear the word "cube", say the drawing of a cube. In what sense can this picture fit or fail to fit a use of the word "cube"? – Perhaps you say: "It's quite simple; – if that picture occurs to me and I point to a triangular prism for instance, and say it is a cube, then this use of the word doesn't fit the picture." – But doesn't it fit? I have purposely so chosen the example that it is quite easy to imagine a *method of projection* according to which the picture does fit after all. – The picture of the cube did indeed *suggest* a certain use to us, but it was possible for me to use it differently.

35. *PI*, § 389. As Judith Genova puts it: "Images, like pictures, still retain a distance from phenomena. They exist within a method of representation and thus are as constructed as any picture. The important point is that they have no privileged position with respect to our thinking and, in fact, are poorer sources of a way of seeing than the pictures embedded in ordinary language" (*Wittgenstein: A Way of Seeing*, London: Routledge, 1995, p. 74). See also her brilliant short paper "Wittgenstein on Thinking: Words or Pictures?" (in: R. Casati – G. White, eds., *Philosophy and the Cognitive Sciences*, Kirchberg am Wechsel: ÖLWG, 1993, pp. 163–167) which, as I have recently realized, is really an anticipation of what I am here trying to say.

Along with this paragraph the editors printed Wittgenstein's remark:

> I see a picture; it represents an old man walking up a steep path leaning on a stick. – How? Might it not have looked just the same if he had been sliding downhill in that position? Perhaps a Martian would describe the picture so. I do not need to explain why *we* do not describe it so.

The first part of this latter remark made an impact on no less a figure than Fodor, who, giving due reference to Wittgenstein, wrote in *The Language of Thought*, way back in 1975: "A picture which corresponds to a man walking up a hill forward corresponds equally, and in the same way, to a man sliding down the hill backward."[36] Two comments. First, as I suggested earlier, and will make it explicit later, still images are, psychologically speaking, but *limiting cases* of dynamic ones. With the development of twentieth-century visual culture, this seems to have become the case with regard to physical pictures, too. I find it difficult to swallow that Wittgenstein, who was a movie addict, and who regularly employed the film metaphor especially in his middle phase, did not make use of the idea of animation when discussing pictorial representation. At any rate, one can confidently assert that an animation showing a man walking up the hill does not look just the same, not even to a Martian, as if he were sliding downhill. Second comment: Fodor omits the Wittgensteinian remark "*we* do not describe it so". It is clear what Wittgenstein had in mind: Pictures belong to our way of life, they are part of our language-games; there are *conventions* guiding the use of a great many of the pictures at our disposal, and with such pictures verbal language does not have to play a mediating role. Wittgenstein took it for granted that the *words* we apply have established, conventional, uses. It is strange that, by contrast, in the case of pictures he should have regarded established conventions the exception rather than the rule. An argument by Søren Kjørup deserves to be quoted here."In most situations", Kjørup wrote, "we understand perfectly well what the uses are to which words are put. – So why should it not be possible to imagine situations in which it is just as evident to which uses pictures are put?" To this he added: "More often than not there is no logical reason why we should not be able to perform a certain illocutionary act with pictures" – that is, *do* something with pictures so that they convey meaning – "but it just so happens that there are no rules for performing the

36. Jerry A. Fodor, "Imagistic Representation", in Ned Block, ed., *Imagery*, Cambridge, Mass.: The MIT Press, 1981, p. 68. This text in Block, ed., is taken from Fodor's *The Language of Thought* (1975). The remark appears again, quite disfigured by then, in Zenon W. Pylyshyn's *Computation and Cognition: Towards a Foundation for Cognitive Science*: "As Wittgenstein points out, the image of a man walking up a hill may look exactly like the image of a man walking backward down a hill; yet, if they were my images, there would be no question of their being indeterminate – I would know what they represented" (Cambridge, Mass.: The MIT Press, 1984, p. 41).

act with pictures, although one might easily be devised. And this is not astonishing", Kjørup continued, "considering that whereas verbal language has been used and refined on the tongue of practically every human being as long as human beings have existed at all, pictures were ... scarce before the invention of picture printing around 1400 A.D."[37] At which point Kjørup ends with a reference to the book by William Ivins we have quoted at length earlier. Incidentally, Kjørup's argument does not mention Wittgenstein – his points of reference are Gombrich, Austin, Searle, and Goodman.

We will come to Goodman in a minute, but let me say first that in the literature on Wittgenstein there certainly are tendencies, too, to establish some kind of *continuity* between his early and later views on picturing.[38] In a devastatingly brilliant, as yet still unpublished paper "Ludwig Wittgenstein: A Case Study in Dyslexia" Hintikka – after pointing out that to dyslexics "a metaphor ... or other nonliteral, shortcut uses of language" might amount to an escape route, and that Wittgenstein's "early account of meaning was a dyslexic's wish fulfillment dream, in that it explained symbolic meaning in terms of pictorial meaning" – stresses that "Wittgenstein's doctrine is as much a propositional interpretation of pictorial meaning as a picture 'theory' of propositional meaning". This seems to me to be a formula which establishes a felicitous connection between the manifest views of the early and the later Wittgenstein. Another strategy is to point out that neither the early nor the later views of Wittgenstein on picturing are as straightforward as they are commonly taken to be. Recall the Tractarian notion of *abbildende Beziehung*, "pictorial relationship",[39] consisting of "the correlation of the picture's elements with things".[40] This "pictorial relationship" has exactly the same function as the later concept of a "method of projection"; the idea of convention is there in the *Tractatus*, too. Nor is the idea of *resemblance* missing from the *Investigations*. "Knowing what someone looks like: being able to call up an image – but also: being able to *mimic* his expression. Need one imagine it in order to mimic it? And isn't mimicking it just as good as imagining it?"[41] What merits attention here is less Wittgenstein's untiring endeavour to relegate mental images to a merely secondary place, but rather that he *does* allot them

37. Søren Kjørup, "George Inness and the Battle at Hastings, or Doing Things With Pictures", *The Monist*, vol. 58, no. 2 (April 1974), pp. 222ff. An excellent supplement to Kjørup's paper is Carolyn Korsmeyer,"Pictorial Assertion", *The Journal of Aesthetics and Art Criticism*, XLIII/3, Spring 1985.
38. So already Anthony Kenny, *Wittgenstein*, Penguin Books, 1973. An excellent recent contribution is Anat Biletzki and David Berlin, "The Logic of Making Pictures", in: R. Casati – G. White. eds., *Philosophy and the Cognitive Sciences*, Kirchberg am Wechsel: ÖLWG, 1993, pp. 47–50. – The discontinuity view, however, is still predominant. In his *Picture Theory* W.J.T Mitchell locates the "philosophical enactment" of what he calls the *pictorial turn* "in the thought of Ludwig Wittgenstein, particularly in the apparent paradox of a philosophical career that began with a 'picture theory' of meaning and ended with the appearance of a kind of iconoclasm, a critique of imagery that led him to renounce his earlier pictorialism ..." (Chicago: The University of Chicago Press, 1994, p. 12 – "Wittgenstein's iconophobia and the general anxiety of linguistic philosophy about visual representation", Mitchell adds, "is ... a sure sign that a pictorial turn is taking place.")
39. Pears – McGuinness translation. Ogden has "representing relation".
40. *Tractatus Logico-Philosophicus*, 2.1514.
41. *PI*, § 450.

some place – the passage just quoted follows only two lines upon the interesting remark "We do not realize that we *calculate*, operate, with words, and in the course of time turn them sometimes into one picture, sometimes into another" – and, even more, that he does indeed allow for a possible *likeness* between pictures and what they depict.[42] The dictum pronounced in *Philosophical Grammar*: "Anything can be a picture of anything", is of only limited validity in *Philosophical Investigations*. This particularly holds for the so-called Part II of *Philosophical Investigations*. Wittgenstein's discussion of *seeing as* would not make sense without the presupposition of *pictures as natural signs*. Recall his introduction, at the beginning of Part II, section xi, of what he describes as a "picture-face". "In some respects I stand towards it", he writes, "as I do towards a human face. A child can talk to picture-men or picture-animals, can treat them as it treats dolls." Of course, as Wittgenstein says some pages later, "custom and upbringing" do play a role in how we see pictures;[43] that role, however, might be in some cases just a slight one. There is a passage discussing the "double cross" – a white cross on a black ground or a black cross on a white ground – where Wittgenstein says: "Those two aspects of the double cross ... might be reported simply by pointing alternately to an isolated white and an isolated black cross. One could quite well imagine this", Wittgenstein adds on an uncharacteristic note, "as a primitive reaction in a child even before it could talk."

The later Wittgenstein's method of explaining philosophical points with the help of diagrams – his *Nachlaß* contains some 1300 of them – would have made no sense if he had really adhered to the position that images do not have an unequivocal meaning unless interpreted verbally. This is the point Andreas Roser makes in his important paper "Are There Autonomous Pictures? Remarks on the Graphic Work of Otto Neurath and Ludwig Wittgenstein", written in the mid-nineties.[44] Roser's main argument, very briefly, is that one could not speak of *different applications of the same picture* if one did not distinguish between the picture and its application. Clearly for any visual object to be treated as a picture presupposes a specific *institutional setting*, and – let me add – for something to be

42. Stressing, also, their *instrumentality*. Compare *PI* § 291: "What we call '*descriptions*' are instruments for particular uses. Think of a machine-drawing, a cross-section, an elevation with measurements, which an engineer has before him. Thinking of a description as a word-picture of the facts has something misleading about it: one tends to think only of such pictures as hang on our walls: which seem simply to portray how a thing looks, what it is like. (These pictures are as it were idle.)"
43. *PI*, Part II, p. 201e.
44. Cf. note 12 above. An earlier version of Roser's paper was read at the conference *Wittgenstein y el Circulo de Viena*, organized by the Universidad de Castilla-La Mancha with the collaboration of the Forschungsstelle und Dokumentationszentrum für Österreichische Philosophie, at Toledo, November 3–5, 1995.

recognized as being of a certain shape or colour at all depends on our specific *neurophysiological makeup*. But this makeup and setting being presupposed, there does indeed exist an autonomy of pictures.

As I tried to show in the foregoing, the view of pictorial representation commonly attributed to the later Wittgenstein is only one of the several approaches actually present in the *Philosophical Investigations*. This might explain why Nelson Goodman, whose argument in *Languages of Art* begins in a way that is definitely reminiscent of § 139 of *Philosophical Investigations*, nowhere mentions Wittgenstein. The Austrian he does mention, and is indeed inspired by, is Ernst Gombrich who in *Art and Illusion* made a convincing case for the inherently conventional nature of pictorial representation. *Art and Illusion* has an endnote referring to the duck-rabbit picture in *Philosophical Investigations*, but I see no reason to believe that Gombrich was in any important way influenced by Wittgenstein. At any rate, Gombrich and Goodman soon parted ways. In *Languages of Art* Goodman already complains of Gombrich not being sufficiently radical – Goodman holds, as Gombrich does not hold, that even *perspective* is a mere convention. By 1981 Gombrich, who in recent years has moved closer to a naturalistic account of images,[45] saw in Goodman but an extreme relativist or conventionalist.[46]

Goodman is, first and foremost, a virtuoso in philosophical analysis; but it is not the case that he does not refer to findings in psychology, ethnology, or indeed art history. However, these references are superficial, spurious, phoney; Goodman does not allot any real theoretical role either to the institutional setting of semantic relations or the psychological bases of visual perception. This is what gives his arguments that peculiar blend of irresistible logical force and at the same time utter incredibility that has baffled his commentators ever since *Languages of Art* was published in 1968. Goodman's thesis, summed up early in *Languages of Art* is: "The plain fact is that a picture, to represent an object, must be a symbol for it, stand for it, refer to it; and that no degree of resemblance is sufficient to establish the requisite relationship of reference. Nor is resemblance necessary for reference; almost anything may stand for almost anything else. A picture that represents – like a passage that describes – an object refers to and, more particularly, *denotes* it. Denotation is the core of representation and is independent of resemblance."[47] It lies in the nature of Goodman's arguments that they typically invite, not careful refutation, but polite rejection. "In the case of classical pictorial representation", Searle wrote in 1974, "objects are represented under their *visual* aspects, and a crucial element in their representation is a visual resemblance between the representation and the thing represented ... I do not wish to imply", he added, "that such notions as resemblance and aspect are unproblematical". However, he concluded, pictorial relationship, "at least within the conventions of classical pictorial representation, relies on resemblance between the picture

45. Cf. Mitchell, *Iconology*, p. 38 and *passim*. – For an excellent discussion of the broader philosophical issue of perspective see Kurt Röttgers, "Perspektive – Raumdarstellungen in Literatur und bildender Kunst", in Kurt Röttgers - Monika Schmitz-Emans, *Perspektive in Literatur und bildender Kunst*, Essen: Verlag DIE BLAUE EULE, 1999.
46. See Gombrich, "Image and Code: Scope and Limits of Conventionalism in Pictorial Representation", in Wendy Steiner, ed., *Image and Code*, Ann Arbor (*University of Michigan Studies in the Humanities*, no. 2) 1981, p. 21.
47. *Languages of Art*, Indianapolis: Bobbs-Merrill, 1968, p. 5.

and the object depicted".[48] Or, as Danto put it some years later:

> a significant degree of match between pictures of x and x must exist, irrespective of the cultural determinants of picture-making. It is this matching which ... gives Professor Goodman's account of pictorial representation its implausibility, as he makes no room for it. Goodman supposes pictorial representation to be more or less exhausted through denotation. But denotation is the most external of semantic concepts, as anything can be used to denotate anything, and if all there were to picturing were denoting, pictures would be as names, and names ... demand associative learning. ... we at least want [a picture] to denote what it does denote *because* of the kind of picture it is, and this then brings in matching properties. Were it not for these the meaning of every picture would have to be explained to us ...[49]

It appears however that Goodman's legacy will have to be subjected, sooner rather than later, to some more minute critical analyses,[50] since it is not without influence on, and causes confusion in, a discipline which today has paramount practical significance. I am referring to cognitive science, and, in particular, the so-called *imagery debate* within cognitive science.

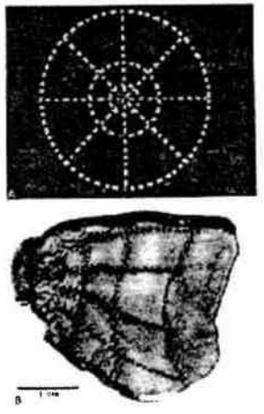
48. John R. Searle, "*Las Meninas* and the Paradoxes of Pictorial Representation", in W.J.T. Mitchell, ed., *The Language of Images*, Chicago: University of Chicago Press, 1974, p. 251.
49. Arthur C. Danto, "Depiction and Description", *Philosophy and Phenomenological Research*, vol. XLIII, no. 1 (Sept. 1982), p. 17. The German translation of this paper is included in the important collection *Was ist ein Bild?* (Gottfried Boehm, ed., München: Wilhelm Fink Verlag, 1994). I am grateful to Karlheinz Lüdeking for having informed me, at the Kirchberg conference in 1995, of the existence of this volume.
50. The first such analysis came, perhaps not surprisingly, from a Wittgensteinian neighbourhood. Richard Wollheim's review of *Languages of Art* (published in *The Journal of Philosophy* in 1970) builds on arguments previously formulated in his book *Art and Its Objects* (New York: Harper & Row, 1968), arguments relying heavily on both Part I and Part II of the *Philosophical Investigations*, on Gombrich, and, to some extent, on Peirce. In the book Wollheim develops Wittgenstein's concept of "seeing as" into a notion of *representation*, stressing that though "the concept of resemblance is notoriously elliptical, or, at any rate, context-dependent", the "attribution of resemblance" does indeed play a role if occurring *within* the "language of representation" (*loc. cit.*, pp. 14–16). In the review "seeing as" leads to "seeing in" – Goodman's convention theory, says Wollheim, is false, since the artist's freedom to employ any picture to represent any object will be limited by what the viewer can *see in* the picture. Wollheim's point was taken up by David Carrier among others, in his contribution to the April 1974 *Monist* issue devoted to *Languages of Art* (as Carrier puts it: "The convention theory is wrong, one wants to say, because it is not true that anything can be seen as representing anything else, even though anything can denote anything else"), or most recently again by Jenefer Robinson in her "*Languages of Art* at the Turn of the Century" (*The Journal of Aesthetics and Art Criticism*, vol. 58, no. 3, Summer 2000: *SYMPOSIUM: The Legacy of Nelson Goodman*). A slight echo of Wollheim can be sensed in Craig Files' excellent paper "Goodman's Rejection of Resemblance" (*British Journal of Aesthetics*, vol. 36, no. 4, Oct. 1996 – the question of "content", writes Files, "is an issue *within* the question of representation"), a paper otherwise very much based on Peirce.

§ 4. The Imagery Debate

There are two topics I would like to discuss in connection with this debate. The first concerns the problem of what *kind* of entities mental images are. The second is about the question of how mental images represent *concepts*. Does not the specificity of images disqualify them from capturing general meanings? I will introduce the first topic by quoting a passage from Stephen Kosslyn, for many years the main protagonist on the image side of the debate. This is how, in his 1994 book *Image and Brain*, he looks back on the early phase of the discussion:

> The issue was not whether people experience mental images. All parties agreed that they do. The issue was not whether propositional representations are sometimes used in cognition. All parties agreed that they are. And the issue was not whether images are solely depictive representations. All parties agreed that for a depiction (a picture or mental image) to have meaning, it must be interpreted in a specific way, which involves a propositional component.[51]

At this point Kosslyn mentions some items from the literature, among them Jerry Fodor's 1975 *The Language of Thought* and, also, Wittgenstein's *Philosophical Investigations*. Clearly, in the imagery debate Wittgenstein stands for the view that pictures without a verbal interpretation cannot carry meaning. Kosslyn then closes the passage by writing: "The issue was whether visual mental images rely on depictive representations (which are in turn interpreted by other processes), or whether they are purely propositional representations." As Kosslyn sees the matter, it was through *neurophysiological* research that a turning-point in the discussion was brought about. By 1982 it became possible to demonstrate that in the cortex there are some visual areas which are *retinotopically mapped*. The neurons in those cortical areas are organized



Pattern stared at by monkey - cortical pattern observed by researchers.

Experiment by Tootell et al. After Kosslyn, *Image and Brain*.

51. Stephen M. Kosslyn, *Image and Brain: The Resolution of the Imagery Debate*, Cambridge, Mass.: The MIT Press, 1994, p. 6. As Barbara Tversky has illuminatingly put it in her talk at the Kirchberg symposium: "The information about environments in long-term memory does not resemble in either structure or format a map such as those developed by cartographers for a variety of purposes. 'Cognitive maps' – that is, whatever mental information is used to make geographic judgments – ... seem to have a variety of formats consisting in part from memory for maps that may have been studied, in part from memory for experiences in environment, in part from memory for descriptions of environments."

to preserve the structure of the retina. "These areas", suggests Kosslyn, "represent information depictively in the most literal sense ... imagery relies on topographically organized regions of cortex, which support depictive representations."[52] Now what Kosslyn does not explain is how we should construe the relation of such cortical patterns to the images we experience as mental contents. It seems to me that the complaint repeatedly raised by Zenon Pylyshyn – the main exponent of the propositionalist side in the debate – namely that the imagist approach lacks a coherent view of its methodological foundations, has been, until very recently, justified.

One retains this impression when looking at the important book *Descartes' Error* by a leading neurophysiologist, Antonio Damasio, published in the same year as Kosslyn's volume.[53] Let me sum up Damasio's stand on mental representation. It is in the form of images, he holds, that the factual knowledge required for reasoning and decision making comes to the mind. Images are not stored as facsimile pictures of things, or events, or words, or sentences. Given the huge quantities of knowledge we acquire across our lives, facsimile storage would pose problems of capacity which would almost certainly be insurmountable. We are all aware that in recalling a face, or an event, we generate not an exact reproduction but rather some sort of re-interpretation, a new version of the original which will in addition evolve over time. On the other hand however we all equally have the sensation that we can indeed conjure up, in our mind's eye, approximations of images we previously experienced. These images images tend to be held in consciousness only fleetingly, and although they may appear to be good replicas, they are often inaccurate or incomplete. Images are the main content of our thoughts. But "hidden behind those images, never or rarely knowable by us", there are numerous processes that guide the generation and deployment of images. "Those processes utilize rules and strategies embodied in dispositional representations. They are *essential* for our thinking but are not a *content* of our thoughts."[54] Damasio, too, attaches high importance to the idea of retinotopical mapping, but is at the same time heir to the work of Frederic Bartlett who in his 1932 classic, *Remembering*, emphasized that mental images have a fundamental role to play in *consciousness*, indeed that consciousness is nothing else but, as he put it, the "turning round" of an organism upon its own "schemata". Now cortical patterns and neurophysiological processes on the one hand, and conscious images on the other, are very different kinds of entities. The methodologically inevitable step is to posit mental images as *theoretical constructs* – as "theoretical entities" exactly in the sense employed by Sellars in "Empiricism and the Philosophy of Mind"[55] – and to treat both the objective and the sub-

52. *Ibid.*, pp. 13 and 19.

53. Damasio, *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: Grosset – Putnam, 1994. The title of the book refers to Descartes' view that the absence of emotions enhances rational thinking. In the course of his work as a neurophysiologist Damasio has come to the conclusion that, on the contrary, without an emotional background no reasoning at all can occur. Hence his dictum: "The traditional views on the nature of rationality [can] not be correct" (p. xi).

54. See *loc. cit.*, pp. 96–108.

55. First published in *Minnesota Studies in the Philosophy of Science*, vol. I: *The Foundations of Science and the Concepts of Psychology and Psychoanalysis*, ed. by Herbert Feigl and Michael Scriven, Minneapolis: University of Minnesota Press, 1956, reprinted in Wilfrid Sellars, *Sci-*

jective sides of the observational data as *empirical correlates* of those constructs. Until very recently this inevitable step could simply not be taken, since the *medium* in which to represent mental images as theoretical constructs was not available. Precisely those dimensions that set images apart from words cannot be described in verbal language; but they can be captured by visual means. The *iconic revolution*, made possible by the graphical capabilities of computer software which barely existed ten or fifteen years ago, now provides us with the instruments of a language in which verbal and visual elements coalesce. Several, juxtaposed, layers of changes have to be taken into consideration here. There is a decreasing dominance of written language and a rise of a new visuality, a process well under way by the 1980s. In his pioneering book *Cognitive Psychology*, published in 1967, Ulric Neisser noted that since *eidetic imagery* – mental imagery of a quasi-sensory vividness and richness of detail – is not uncommon in young children, but very rare among adults (namely *American* adults), that capacity must somehow diminish with age. "Some visual factor connected with literacy", Neisser remarked, "may be responsible".[56] Recall the related points made by Russell and Price. We might hypothesize that the ability to *have* images is, today, again on the ascent – this is what I regard as a first layer of change. Secondly, people are becoming *familiar* with pictures, are acquiring a rich experience of dealing with pictures, to an extent unprecedented throughout written history. And thirdly, to repeat, there is the change connected to the use of today's computers: the ease with which one can *produce* pictures, the increasing everyday possibility to *communicate* via pictures.

It is in the work of Lawrence Barsalou that the imagery debate has taken a first step towards the methodological clarity the iconic revolution makes possible.[57] In his paper

ence, Perception and Reality, London: Routledge & Kegan Paul, 1963. One of the first contributions to the present phase of the imagery debate, Allan Paivio's *Imagery and Verbal Processes* (New York: Holt, Rinehart and Winston, 1971), represents an entirely clear methodological position here. Both "mental images and mental words", writes Paivio, belong to the order of "postulated processes", both are "theoretical constructs", "inferential concepts", which can have "functional significance" only to the extent that each "can be differentiated from other concepts theoretically", and to the extent that "these distinctive theoretical properties are open to empirical test". Paivio's question is "whether or not it is necessary, or at least useful, to postulate both kinds of symbolic processes, nonverbal as well as verbal, to account for effects that have been observed in a variety of situations". He contrasts his own methodology with "the classical approach to imagery" in which "the term image was used to refer to consciously-experienced mental processes". (*Imagery and Verbal Processes*, pp. 6–11.) This contrast became blurred again in the later discussions.

56. New York: Appleton-Century-Crofts, pp. 149f. Note however that, as Barbara Tversky has pointed out in the discussion at Kirchberg, Neisser's evidence has been contested by a study by Haber who found no decline with age of eidetikers.

57. In his method of iconic representations Barsalou is heavily indebted to Ronald W. Langacker's 1986 essay "An Introduction to Cognitive Grammar" (*Cognitive Science* 10). Langacker's focus, however, is on linguistic meaning rather than on mental imagery. By the mid-eighties the new graphic capabilities of computers have definitely begun to make themselves felt. In their book *Understanding Computers and Cognition* Winograd and Flores could already write about "the appeal of computers like the Apple MacIntosh (and its predecessor the Xerox Star)", an appeal being due to encompassing both "text and graphic manipulation" (Reading, Mass.: Addison-Wesley, 1986, p. 165).

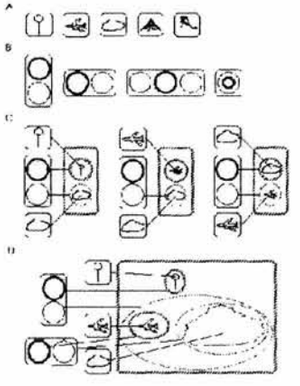
"Perceptual Symbol Systems", published last year in *Behavioral and Brain Sciences*, Barsalou outlines an approach to the problem of mental images that is indebted to a wide array of sources, among them certain arguments set forth by Searle in his 1980 "Minds, Brains, and Programs". I have no space here for giving a complete summary of Barsalou's approach; I single out just a few main points.

Barsalou's position is that cognition is inherently perceptual, and that perceptual memories can function symbolically, which means: 1. they can stand for referents in the world, and 2. they can allow for symbol manipulation. However, as Barsalou underlines: "Perceptual symbols are *not* like physical pictures; nor are they mental images or any other form of conscious subjective experience. As natural and traditional as it is to think of perceptual symbols in these ways, this is not the form they take here. Instead, they are records of the neural states that underlie perception." Perceptual symbols combine, and form systems; and perceptual symbol combinations are, as Barsalou puts it, productive, they represent not holistically but componentially. Barsalou introduces iconic conventions to denote perceptual symbols and their combinations, but stresses, again, that his diagrams "should *not* be viewed as literally representing pictures or conscious images. Instead, these theoretical illustrations *stand for* configurations of neurons that become active in representing the physical information conveyed in these drawings."

Lawrence Barsalou
illustrating the productivity of
perceptual symbol systems:

BALLOON, JET, CLOUD,
HILL, KITE

ABOVE/BELOW,
LEFT-OF



"An example of how perceptual symbols for object categories (A) and spatial relations (B) implement productivity through combinatorial (C) and recursive (D) processing. Boxes with thin solid lines represent simulators; boxes with thick dashed lines represent simulations." The term 'simulator' (suggested by Stevan Harnad) is a replacement for 'simulation competence' in Barsalou's previous articles.

Perceptual symbols do not just represent, or refer to, classes of objects; a system of such symbols is also suited to express *propositions*. Barsalou presents diagrams expressing that some perceived individual object *belongs* to a certain class, or that *it is true that* some perceived objects stand in a certain *relation* to each other, repeatedly emphasizing that "such drawings are theoretical notations that should not be viewed as literal images". Now Barsalou's reason for these reiterated affirmations that neither perceptual symbols, nor indeed the diagrams he uses, do in any actual sense *depict*, seems to be that he accepts Goodman's position. He refers to Goodman when he writes: "the degree to which a symbol's content resembles its referent is neither sufficient nor necessary for establishing reference". But at the same time he cannot quite do without the idea of real pictoriality. As he puts it: "variations in the form of a perceptual symbol can have semantic implica-

tions". Or again: "As the content of a symbol varies, its reference may vary as well." He tries to have it both ways when he says: "Just because the content of a perceptual representation is not the only factor in establishing reference, it does not follow that a perceptual representation cannot have reference and thereby not function symbolically." In the end he just gives up. "The conditions under which analogical reference holds", he writes, "remain to be determined." But Barsalou's tendency to shun any reliance on resemblance has the undesirable consequence that he actually has to ban imagery from consciousness. His "drawings", he writes, "are used *only* for ease of illustration", and "they stand for unconscious neural representations". It appears that, with Goodman in the background, the theoretical constructs Barsalou introduces cannot quite fulfil their role. The way out of this predicament is the way back to a philosopher Barsalou regularly refers to, the philosopher to whom he really owes a fundamental debt: H.H. Price. Price is the philosopher, too, in whose work we find the most elaborate attempt to deal with the issue of generic images.

§ 5. Generic Images

Locke's problem was that no mental image of a triangle can stand for the general concept of a triangle, since any image would have to be a *determinate* kind of figure. Daniel Dennett seems to have run into a similar problem when he discovered, way back in 1969, that whereas one can speak of a tiger having numerous stripes, one cannot have a mental image of a tiger without experiencing a *particular* number of stripes. Hence imagining a generic tiger cannot amount to having the mental image of a tiger.[58] Fodor, in *The Language of Thought*, attempts to refute Dennett. He points out – correctly, I believe – that a mental image can be *blurred*, or can be "a sort of a transient stick figure"; and then closes the argument – falsely, I think – with the observation that even a blurred, transient, or simply bad, image can have the proper reference, *if* it is connected with one's intentions in the right way.[59] In an exchange which must have been written at the time Fodor's book was published, Elliott Sober and Robert Howell had a rather convoluted discussion in *Synthese*, approximating, between them, the conclusion that pictures need not be determinate. Sober asserted – against Berkeley – that "there *could* be a picture of something triangular which leaves indeterminate the kind of triangularity the thing has", if, say, "the object is pictured as being in the distance, at an angle difficult to determine, and partially obscured by fog"; while Howell, wrestling with one of Sober's examples, made the insightful comment that if we imagine "a picture showing a line of fire engines fading off into far distance", the little, quite blurred, specks towards the end of the line will still be taken as *pictures* of fire engines.[60] These discussions from the 1970s however, though useful, were altogether inferior to the analysis provided by Price in his *Thinking and Experience*, published in 1953 – the same year Wittgenstein's *Philosophical Investigations* came out.

58. See ch. 2 of his *Content and Consciousness*, repr. in Ned Block, ed., *Imagery*.
59. Jerry A. Fodor, "Imagistic Representation", in Ned Block, ed., *Imagery*.
60. Elliott Sober, "Mental Representations", and Robert Howell,"Ordinary Pictures, Mental Representations, and Logical Forms", both in *Synthese* 33 (1976).

Price, Wykeham Professor of Logic in the University of Oxford, nowhere even mentions Wittgenstein in his book, but it is clear that the latter is indeed his *bête noir*. My impression is that the significance of *Thinking and Experience* was never really appreciated, and that the book in fact became submerged in the torrent of Wittgensteinianism in the 1950s and 60s. It is a brilliant book, deserving to be discovered at long last. Here I can only mention those of its main results which have a direct bearing on the argument of my talk. Price insists that some of us do *use* images in our thinking.[61] "Modern philosophers", he writes, "are never tired of telling us that mental images are not at all like pictures. But they are."[62] Images, says Price, have a superiority over words, in that "they come *nearer* than words do to being instances of the concepts brought to mind by means of them." They are in this sense *quasi-instantiative particulars*, "whereas words ... are completely non-instantiative particulars. Thus when we think in images, thinking in absence comes much nearer to perceiving in presence than verbal thinking can."[63] However, Price also insists that although mental images are quasi-instantiative particulars, they are not the only ones. "Models, diagrams, pictures drawn publicly in the light of day with nothing 'mental' about them, ... public cinematographic reproductions ... all these entities and occurrences have the same quasi-instantiative function as images have."[64]

Now the quasi-instantiative function of both mental images and physical replicas clearly relies on *resemblance*. Price does not believe that the notion of resemblance is unproblematic. In fact he discusses three types of problems: that of *too little* resemblance, that of resembling *too many* things, and that of resembling one thing *too closely*. The problem of too little resemblance has really two aspects to it, a trivial one and and an intricate one. When an image is too faint, or a replica too imperfect, we might just say that they are *useless*, as there are also words which are so vague as to be useless. This is, then, a trivial aspect of too little resemblance.[65] However, a hazy image or a bad replica might also have the *advantage* of not being bound too closely to any particular instance of the concept instantiated. This is an intricate matter, leading to the issue of generic images. The problem of resembling too many things is a problem of *ambiguity*. One's image of a crocodile might be a very clear one, but that image, points out Price, will then resemble not only a crocodile, but also a lizard, a reptile, an animal, and so on. The image-symbol, writes Price, seems to have "a kind of systematic ambiguity within the determinable-determinate hierarchy it refers too".[66] But there is also non-systematic ambiguity. The image of a crocodile resembles anything with roughly the same colour, and anything with roughly the same shape. As Price puts it: "any image symbol, or indeed any replica whether physical or mental, is bound to be ambigious. ... it cannot help having too many resemblances. In this respect ... the particulars which symbolize by resemblance are greatly inferior to words and other non-resemblant symbols."[67] There is a solution to this predicament. "When I think about some object or class of objects in an imagy manner",

61. *Thinking and Experience*, p. 235.
62. *Ibid.*, p. 249.
63. *Ibid.*, pp. 254f.
64. *Ibid.*, p. 256.
65. *Ibid.*, pp. 266f.
66. *Ibid.*, p. 268.
67. *Ibid.*, p. 270.

says Price, "I am not restricted to using just one single image. I might use a series of different images. Again, the image which I use need not be static. It might be, as it were, a working model, cinematographic rather than static."[68] This is a momentous, far-reaching insight. And it applies to the difficulty posed by systematic ambiguity, too. What determines, Price asks, whether a crocodile image means *crocodile*, or *reptile*, or *organism* in general? "There may be no difference", he answers, "if we consider just one single image, especially if it is a static image. But there is a great difference if we consider what *other* images we produce, or have a tendency to produce, along with or after this one."[69] What Price here does, then, is to apply the *context principle* to pictorial meaning. And it is significant that in this connection he refers to Hume's formula, in the section "Of Abstract Ideas" of Book 1 in the *Treatise*, according to which we have ideas "not really and in fact present to the mind, but only in power", ideas we do not "draw ... all out distinctly in the imagination, but keep ourselves in a readiness to survey any of them, as we may be prompted by a present design or necessity".

A different problem is posed by images resembling one thing too closely. The more closely a mental image resembles a particular entity, suggests Price, the less closely it will resemble many others in the same class. Images should certainly be fit to fulfil the role of standing for *concepts*, and from this point of view "it would almost seem that a 'bad' image – schematic, sketchy, lacking in detail – is better than a 'good' one".[70] Price points out that "the problem of 'too much resemblance' still arises when we consider the relation between an image and an individual object".[71] The more an image resembles, say, the front view of some particular thing, the less it will resemble its side view. As Price sums it up: "It seems as if an image must have some at least of the properties of a general symbol if it is to represent even an individual object adequately."[72] There are two kinds of images Price draws attention to in this connection. First, generic images of a Galtonian sort, analogous to composite photographs, which, as Price puts it, "symbolize by typical resemblance". Images of this kind, Price believes, "do occur, and can be used for thinking of classes whose members differ from each other, provided these differences are not too great".[73] Secondly, and more importantly, there are what Price calls "inchoate images",[74]

68. *Ibid.*, p. 272.
69. *Ibid.*, p. 273.
70. *Ibid.*, p. 275.
71. *Ibid.*, p. 282.
72. *Ibid.*, p. 283.
73. *Ibid.*, p. 293.
74. "Do we have images which are as it were inchoate entities? I confess that I cannot confidently answer 'No', as Berkeley and Hume did. ... Ordinary people, who are neither philosophers nor psychologists, sometimes describe their images as 'vague'. Indeed, anyone who has images at all is strongly tempted to describe some of them in this way. Yet a philosophical purist might object that it makes no sense to call an actual particular vague. When a word is called vague, for example the word 'bald', the vagueness belongs not to the sound or mark itself (there is nothing vague about that) but rather to the symbolic function it performs. We cannot draw a sharp line between the things to which it applies and the things to which it does not. But when an image is called vague, it is not vagueness of meaning that is referred to, but something in the nature of the image itself, an intrinsic or internal vagueness. ... I would suggest that the word 'vague', in this usage, is just an untechnical and perhaps misleading equivalent for the terms 'inchoate', 'incompletely determinate', ... and likewise for Locke's term 'imperfect'", *ibid.*, pp. 288f.

of a "vague", "evanescent", or "fleeting" character.[75] The doctrine of general images is defensible, *if* those images are conceived as "not fully determinate particulars". As Price sums it up: "It would appear that such incompletely determinate particulars do occur in image thinking, however we choose to describe them. And if indeed they do, they seem well fitted to serve as general symbols ..."[76] But here Price adds that if there *are* incompletely determinate images, "the analogy between images and physical replicas ... to that extent breaks down".[77] In the domain of *public communication* the problem of generic pictorial meaning is left unresolved by Price. It can be resolved by calling attention to the means of *pictorial conventions*. Those means however were much less available in Price's times than they are today. We are back at the topic of the iconic revolution.

We are back, also, at Barsalou. We are back, strengthened in our belief that 1. *resemblance* does indeed play a role in mental imagery, 2. *conscious* mental images ought to be taken note of in theories of mind, 3. images can be indeterminate, schematic, and hence *generic*, and 4. the *ambiguity* of images can be overcome once the ideas of context and, in particular, *sequentiality* are made use of. I think Barsalou has no entirely clear picture of where actually he stands in relation to Price, whom he often refers to;[78] but he certainly is inspired by him. My impression is that even Barsalou's contention that *logical terms* can be made sense of in terms of perceptual symbols goes back to Price who, albeit only in passing, did entertain the possibility of there being, in the realm of mental images, equivalents for "logical words".[79]

§ 6. Conclusion

Let me sum up. In this talk I have attempted to show that, due mainly to advances in computer software, pictures are today becoming a vehicle for communicating rational thought. As a direct result of this, it is increasingly possible to introduce, namely *form pictures of*, theoretical entities which can explain how the subjective experience of mental imagery hangs together with neurophysiological facts. The brain works neither with pictures, nor with propositions. It works with clusters of functionally connected neurons. But how can we best convey an idea of those functional connections? Using pictures extensively in our explanations is both more natural and theoretically more fruitful than

75. *Ibid.*, p. 290.
76. *Ibid.*, pp. 292f.
77. *Ibid.*, p. 293.
78. A passage I find very instructive in that it shows in a nutshell Barsalou's difficulties in coming to terms with contradicting literary influences: "There is good reason to believe that perceptual representations can and do have intentionality. Pictures, physical replicas, and movies often refer clearly to specific entities and events in the world ([cf.,] e.g., Goodman, [*Languages of Art*]; Price, [*Thinking and Experience*])." At which point follows the passage I have already quoted: "Just because the content of a perceptual representation is not the only factor in establishing reference, it does not follow that a perceptual representation cannot have reference and thereby not function symbolically."
79. *Thinking and Experience*, pp. 104 and 297.

talking merely in terms of formal strings of abstract symbols. Until quite recently, this avenue was simply not open when it came to philosophical communication. Now, however, it even becomes possible to give a new interpretation of the history of the philosophy of mind, explaining why the common-sense belief that we, really, *think in images*, never received an adequate theoretical formulation.

Thus is a whole cloud of philosophy condensed into a drop of communication technology. To paraphrase Wittgenstein in this way is tantamount to criticizing him. However, all along I have tried to present my case in such a manner that some basic ideas of the later Wittgenstein, ideas I still regard to be correct, should not be jeopardized. I am referring to ideas like: meaning and intention always depend on context; thinking is an activity involving not just our brain, and not just our body, but also factors, structures, and institutions *external* to the human organism. These ideas are, I think, wonderfully epitomized by Wittgenstein's dictum: "If God had looked into our minds he would not have been able to see there whom we were speaking of."[80]

80. *PU*, Part II, p. 217e.

The Irrationality of Religion
A Plea for Atheism

HERMAN PHILIPSE

1. Introduction

In this paper I attempt to substantiate the thesis that the core-beliefs of religions are irrational. These core-beliefs are the monotheist contention that there is one God or the polytheist opinion that there are a number of different gods. Outside mathematics, the word 'irrational' may signify two different things. Either it means that a sentient being is not endowed with reason, for instance if one speaks of 'irrational animals' such as slugs. Or it means that a belief or an action is contrary to reason, that is, unreasonable, utterly illogical, or absurd. I claim that all religious core-beliefs are irrational in this second sense. And of course, irrationality should be avoided.

It will be objected to my thesis that beliefs cannot be accused of being unreasonable unless they are situated within the province of reason. Could one not argue that religious beliefs are not located within this province because, as Pascal said, 'the heart has reasons which reason does not grasp'? According to some religious authors, the domain of reason is somehow limited, and faith must be situated entirely, or in part, beyond the limits of human reason. I shall argue that even if faith transcends reason in this manner, the core-beliefs of religions are unreasonable.

2. Strategy

While believers are always partial or biased in religious matters because they prefer their own religion to the others – even a syncretistic religion is a specific faith that is logically incompatible with other religions –, the atheist must be epistemically impartial in that he rejects religious favouritism. His arguments should refute each and every religious existence claim, and they have to hold against all interpretations of each religion, orthodox and more liberal ones. Furthermore, atheism as such is restricted to the epistemic aspect of religions: the atheist rejects the idea that a god or that gods exist, hence he might appreciate from an aesthetic or moral point of view many different religious forms of life, taking an anthropological stance. This neutrality with regard to specific religions puts the atheist in a difficult argumentative position. The number of religions is large, and the set of actual and possible interpretations of each of them is perhaps an infinite one. How will the atheist be able to argue against all religious existence claims at once?

The most promising strategy is to proceed by way of dilemmas. If the atheist is able to construe a dilemma concerning all religions, which exhausts the entire field of religious possibilities because its two horns are each other's contradictories, and if the atheist is

B. Brogaard, B. Smith (Hg.), *Rationality and Irrationality / Rationalität und Irrationalität*, S. 267–272.

able to show that each horn gives rise to atheism, he will have won the battle. I shall briefly sketch such an argumentative strategy, which I have further developed elsewhere. It consists of a series of dilemmas, starting with the overarching dilemma that religious faith either (A) transcends reason or (B) finds itself within the province of reason. Let me call this the dilemma of faith and reason. I shall argue that the horns of the dilemma of faith and reason imply atheism.

Before I develop my argument, two clarifications will be useful. First, the dilemma might be applied both to large clusters of religious tenets and to individual claims to religious truth. For instance, in the Scholastic tradition some elements of Christian faith, such as the claim that there is an infinite god, were considered to be sustainable by reason, whereas others, such as the dogma of the monotheist unity of Father, Son, and Holy Spirit, were often thought to transcend reason and to be known by revelation only. It is up to believers to determine which religious tenet they want to put on which horn of the dilemma. Incidentally, since the dilemma applies to each and every religious tenet, its two horns are contradictories instead of contraries.

Second, the atheist is free to define the notion of reason as he pleases, provided that he uses his notion consistently. For the sake of argument, I define 'reason' as the methods of empirical research and critical discursive thought. It is no objection to my overarching dilemma that the term 'reason' might be defined differently, for instance as Hegelian *Vernunft* as opposed to Hegelian *Verstand*. The only legitimate objections are either that at least one of the horns does not lead to atheism ("grasping the dilemma by the horns"), or that the two horns do not exhaust the field of possible religious positions ("escaping between the horns"), or that the believer is able to launch a counter dilemma which destroys atheism ("rebuttal").

3. Faith Beyond Reason

Let us start, then, by discussing the first horn (A) of my overarching dilemma, the horn according to which faith transcends reason. In other words, the first horn says that religious belief is a-rational. This idea is old and venerable. Allegedly, the object of faith is too sublime to be grasped by a down-to-earth capacity such as human reason. But accepting this horn leads one inexorably to an atheist conclusion of a peculiar kind, as the following reflections will show. Hence, if religious belief is a-rational, it is irrational.

Everyone should agree that religious belief is impossible without a propositional content, for one cannot believe without believing that some proposition is true. Indeed, believing is always believing that p, and believing that p is accepting as true that p (these are observations on the logical grammar of 'to believe'). The minimal content of a religious belief is the proposition that gods, or God, exist or exists, and according to my precising definition of 'religion' there simply is no religion without acceptance of such a proposition. Propositions are expressed by meaningful declarative sentences. Furthermore, declarative sentences cannot be meaningful if one of the words *used* in the sentence is meaningless. Hence there cannot be religious belief unless the word 'god' or 'God' has been assigned a meaning.

It is not up to the atheist to define the proper name 'God' or the common noun 'god',

except in the minimal sense of a supernatural entity. The atheist leaves this task to the religious believer. From the impartial point of view of the atheist, believers have an immense room for choice here, since innumerable descriptive definitions of gods have been provided in the history of mankind, and believers might conjure up indefinitely many new definitions. As we will see, however, the set of all possible definitions is exhaustively divided into two subsets by the dilemma that either faith transcends reason or faith is located within the province of reason.

If one opts for the horn that faith transcends reason, the descriptive definition of 'god' or 'God' which gives meaning to the religious proposition 'God exists' has to meet specific requirements, supposing at least that one wants to believe that God (or gods) exist(s). For if faith transcends reason in the sense defined above, that is, empirical inquiry and discursive thought, it must be a priori impossible that a religious core-belief be refuted by empirical data. In other words, the definition of the word 'god' has to meet the postulate of empirical irrefutability (apart from the logical requirement of consistency) in the following sense: the belief that God, or a god, as defined by the definiens, exists in fact, must be immune to all possible empirical refutations.

The postulate of empirical irrefutability implies another postulate, which I call the postulate of factual emptiness. Although there are events that according to current science cannot be investigated by direct empirical research, such as the events of which special relativity theory says that they are outside our 'light cone', one cannot exclude a priori that these events will ever be open to empirical investigation of a more indirect kind. Moreover, the idea that there are facts or events which are a priori outside the empirical domain presupposes that there may be true synthetic a priori propositions, a presupposition that is now generally rejected by philosophers. Consequently, as there can be no facts of which one might guarantee a priori that they are not open to empirical investigation, the postulate of empirical irrefutability implies the postulate of factual emptiness.

Accordingly, the believer who claims that faith transcends reason has to give a descriptive definition of 'God' or 'god' such that the proposition 'God exists' is devoid of factual content. This postulate of factual emptiness has been endorsed by religious philosophers of the twentieth century such as the early Wittgenstein and the later Heidegger. Defining the world as the totality of facts, Wittgenstein claimed that 'God does not reveal himself *in* the world' (*Tractatus*, 6.432), and Heidegger held that *Seyn* radically transcends the world of *Seiendes*.

But clearly, the postulate of factual emptiness destroys the possibility of defining 'God' or 'god' altogether, because a descriptive definition of 'x' such that 'x exists' is devoid of factual content is impossible: a fact is precisely what obtains if a descriptive statement is true. We now see the devastating implications of the first horn (A) of our overarching dilemma. If faith transcends reason, we cannot give meaning to the word 'god' or 'God'. Consequently, the phrase 'God exists' is meaningless, and the claims that God exists (faith), that we do not know whether God exists (agnosticism), and that God does not exist (traditional atheism) are also meaningless. I call this implication *semantical atheism*, for if one cannot give meaning to the thesis that God exists, religious belief is impossible.

We must conclude that the first horn of the dilemma is self-refuting. If the very idea that faith transcends reason implies, by a chain of arguments, that faith is impossible if it

transcends reason, faith cannot transcend reason. I said that my overarching dilemma divides the set of all possible definitions of 'god' into two subsets. Clearly one of these subsets equals the null class, for no definition of 'god' can satisfy the requirement of factual emptiness.

As a consequence, the would-be believer who claims that faith transcends reason, is landed in a second dilemma. Either he continues using the word 'god' without having given meaning to it, which, I suspect, is the case of many popular religious authors, or, if he has provided a descriptive definition of 'god' or 'God', this definition has descriptive content. In the latter case, his claim that God exists will have factual implications. Hence it is in principle refutable by empirical research or discursive argument and the believer has not succeeded in transcending reason. This brings me to the second horn of the overarching dilemma, the horn that faith is located within the province of reason.

4. Faith Within the Province of Reason

The believer who locates his faith within the province of reason is confronted by a great number of dilemmas that destroy his position, for now his faith is answerable to reason. It will suffice here to point out three interconnected dilemmas from which the believer cannot escape. If faith is located within the domain of reason, one should raise the question as to what explains the fact that the believer has faith. This question triggers a first dilemma: either (C) one gives a religious explanation or (D) one gives a secular explanation. Assuming that each of these explanations points to a cluster of causes of faith that is sufficient to explain the presence of faith, the religious and the secular explanation exclude each other because of Occam's razor. Even believers who reject natural theology, holding that reason cannot bring us before God, might locate their faith within the province of reason if (a) they provide us with a religious explanation of their faith and (b) they are prepared to bring their religious explanation into competition with secular explanations and to adjudicate between these competing explanations by using accepted criteria of theory choice.

For instance, believers might explain the fact that they have faith by saying that God's grace bestowed faith upon them, and that this grace is a sufficient condition for faith. Such is the traditional Christian explanation given by Paul, Luther, and many others. Clearly this explanation not only explains the presence of faith in believers but also justifies it: if God caused faith in us, God must exist, and the belief that He exists is true. For this reason we might call religious explanations of faith self-justifying. The question is, however, whether such a religious explanation is acceptable according to the usual criteria of theory choice. Should one not prefer a secular explanation? Secular explanations all belong to the class to which Freud's theory of projection belongs. They typically start from the assumption that the belief that God or gods exist(s) is not true and try to explain the fact that the believer has the illusion that God or gods exist(s) by pointing to psychological, sociological, or other secular mechanisms. Clearly, then, the religious and the secular explanations are mutually incompatible, because they contradict each other.

A second dilemma shows that according to accepted criteria of theory choice, one should always prefer an explanation of the secular type to religious explanations of the

presence of faith in believers. This dilemma starts from the fact that there is a plurality of religions and that the religious contents of these religions contradict each other at many points. For example, monotheistic faith contradicts polytheistic beliefs. The dilemma arises for the believer who wants to provide a religious and self-justifying explanation of his own faith. Either (E) this believer provides a religious and self-justifying explanation for his own faith only, explaining the faith of other religions by a theory of projection. But this is an illegitimate move. It is a case of *special pleading*, unless the believer is able to argue convincingly that his own faith is true and the beliefs of competing religions are false, which is unlikely. Or (F) the believer chooses to explain the faith of all religions by means of a religious explanation. However, this second horn triggers a third dilemma: which religious explanation should he prefer?

Either (G) the believer tries to explain the faith of all religions by supposing that the self-justifying explanation of *his own* faith also explains the faith of *other* religions even though these other religions may contradict his own. This is the theory of the Catholic Church, which claims that Catholic faith is absolutely true and is caused by the Catholic God in Catholic believers. The Catholic God would also have caused (or at least "permitted") Hindu beliefs in Hindu believers and polytheist Germanic beliefs in the Germans of the Edda epoch. Why would the veracious Catholic God do such a weird thing? Why would He cause religious beliefs in non-Catholics, beliefs that must be false to the extent that they are incompatible with the absolute truth of Catholicism? The Catholic solution to this embarrassing problem is that the Christian truth is hidden in all other religions, and that believers of these other religions are "on the way" to the Catholic Truth even though they are not quite ready to receive Christian grace. This clearly is an *ad hoc* solution which shipwrecks the attempt to explain all religions by the self-justifying explanation of one's own faith.

Should one then (H) try to explain each faith by supposing that the self-justifying explanation of each and every religion is true? That is, should one suppose that the Catholic faith is to be explained by claiming that the Catholic God caused this faith in Catholics, and that the Hindu faith is to be explained by supposing that each of the innumerably many Hindu gods caused faith in him- or herself in Hindu believers? This possibility is ruled out by the fact that these explanations contradict each other: according to Christian faith, there is only one God. We must conclude that the attempt to explain faith by religious explanations runs into insuperable difficulties. Only a secular explanation is able to account for *all* occurrences of religious faith. Irrespective of their individual scientific credentials, then, secular explanations must be preferred, because they are a priori more empirically adequate than their religious rivals. And because secular explanations of religion start from the assumption that each religious faith is false, those who want to advance in the endeavour of explaining the phenomena of faith must be professional atheists.

I conclude that the overarching dilemma destroys all possibilities for faith. If faith transcends reason, semantical atheism is the result. If faith is located within the province of reason, we must all become traditional atheists. This conclusion substantiates the thesis of the irrationality of religion.

5. Discussion

How is the atheist to discuss religion with the faithful? It is wise to start by saying nothing at all. But if a believer advances the claim that God or a god exists, or engages in a language game such as prayer which presupposes that there is a god, the atheist might ask the believer by which descriptive definition the latter gives meaning to the word 'god'. Typically, such a definition, conjoined to the existence claim, will have factual implications, and the atheist might point out the dilemmas which the believer has to confront (B-H). However, if the believer reacts to these dilemmas by eliminating the factual implications from his definition, the atheist will show that by this move the believer has destroyed the content of his existence claim and that, *eo ipso*, he has ceased to be a believer (A).

A meticulous analysis of religious language games in contemporary Western culture will reveal that believers keep oscillating between these two options and refuse to choose consistently, because they are dimly aware that every consistent choice will shipwreck their religion. This is what we should expect. For religion is the product of man's longing for there being more to life than in fact there is. As soon as one tries to formulate this 'more' in meaningful propositions, one perceives that these propositions are very probably false.

I have called this argument for the irrationality of religious belief a 'plea for atheism', and so it is. One might object that it would be more prudent to become an agnostic. Is atheism not as irrational as religious belief? Indeed, is it not a kind of dogmatic belief itself? But it is easy to see that the overarching dilemma of faith and reason destroys agnosticism as it destroys faith. Agnosticism is the position that we should refrain both from believing that there is a god and from believing that there is no god, because the arguments in favour of one of these positions are not stronger than those in favour of the other. Accordingly, agnosticism does not make sense unless the phrase 'God exists' makes sense. But as we saw, this phrase is meaningless if (A) faith transcends reason. And if (B) faith is located within the domain of reason, the arguments for atheism are stronger than those in favour of a particular religious belief.

Philosophy in Finland – Analytic and Post-analytic

SAMI PIHLSTRÖM

1. Introduction

The history of Finnish analytic philosophy provides an interesting example of the "geography of philosophical reason". The analytic tradition was introduced in Finland by Eino Kaila, who had close contacts with the Vienna Circle in the late 1920s and the early 1930s. Kaila wrote his main works in German – and never became an internationally reputable analytic philosopher, since (as is well known) the center of gravity of analytic philosophy shifted from the German-speaking world to the United States and Great Britain after World War II. In any event, Kaila's pupils (G.H. von Wright, Oiva Ketonen, and Erik Stenius, among others) and von Wright's pupils (especially Jaakko Hintikka) established in Finland the Anglo-American analytic style of philosophizing which has later been continued by such logicians and philosophers of science as Juhani Pietarinen, Risto Hilpinen, Raimo Tuomela, and Ilkka Niiniluoto (and, again, by some of their pupils). Because of their work, Finnish philosophy is an integral part of an international philosophical tradition whose *lingua franca* is English. Thus, it is largely because of Kaila and his followers that the analytic tradition was established as the mainstream philosophical movement in Finland. Analogous developments have taken place in other Scandinavian countries, especially Sweden and Norway.

Analytic philosophy, in Finland and elsewhere, might be interpreted as an attempt to examine the notion of rationality by philosophical and conceptual means. What I mean by this vague idea is that the rules of logic, according to many analytic philosophers, provide us with norms of pure scientific reason and enable us to formulate our thoughts and arguments as clearly and responsibly as possible. The methodological program advocated by analytic philosophers usually consists in the application of logical tools in various philosophical subject matters. Even when no technical formalizations are used, analytic philosophers attempt to do their problematizing, explicating, and argumentative work as carefully and soundly (in a logical sense) as they can. Thus, something that may be called *philosophical analysis* is seen as the fundamental method to be employed more broadly than mere logical analysis.

Rather than explaining what analytic philosophy (or philosophical analysis) is – to an audience for whom such explanations are hardly needed – I want to discuss Finnish analytic philosophy through some hopefully illuminating examples. It is also my hope that this discussion will throw some light on certain peculiar features of the analytic tradition, and its future, outside Finland, too. I am not, so to say, interested in the history of Finnish philosophy for its own sake. Instead, I believe that complex and general problems, such as the nature and development of the analytic tradition, or the importance of the notion of rationality in that tradition, can be fruitfully approached through case studies.

It goes naturally with this attitude to remind analytic philosophers of the fact that phi-

B. Brogaard, B. Smith (Hg.), *Rationality and Irrationality / Rationalität und Irrationalität*, S. 273–283.

losophy is not only done in an international research community but also, at the same time, within the various national traditions whose presuppositions and implicit commitments determine, at least to some extent, how we professional philosophers engage in our properly academic, internationally published research. This is why geographies of philosophical reason may be of genuinely philosophical (and not merely historical) interest. Especially Finnish philosophers, like all philosophers (or academic researchers more generally) representing a small nation and a small language, should be concerned with their national tradition – at least once in a while, at least a little bit. Needless to say, this concern must not lead to parochialism or to second-rate work written only in one's mother tongue instead of the language used by the international research community (i.e., English). One can, though perhaps not easily, combine high-level, intellectually responsible academic philosophy, primarily directed to one's international colleagues, with more popular philosophizing (sometimes even about the same topics), directed to the more general well-educated audience who reads philosophical books in its mother tongue, such as (say) Finnish.

There may also be philosophical reasons for not forgetting one's mother tongue in philosophical writing. If one is impressed by Wittgenstein's later thought, as several analytic philosophers of course are, one may insist on the need to take seriously the form(s) of life and the corresponding language-games that one has primarily been acquainted with as a child in one's early natural and cultural environment, i.e., the conceptual and practical frameworks that have formed the background of one's maturation as a human being. Indeed, using one's mother tongue when philosophizing may be a necessary condition for the possibility of using another language, such as English, as a philosophical working language (a transcendental condition, if that phrase is allowed). A Finnish philosopher may, then, quite legitimately arrive at the conclusion that some of her or his deepest problems and convictions regarding matters of vital human importance – philosophical problems and convictions that may nevertheless reflect ordinary "non-philosophical" experiences in life – can only be adequately expressed in Finnish. This attitude need not conflict with the recognition that serious academic work must be published in an international language. On the contrary, when a Finnish philosopher becomes a member of the international philosophical research community, she or he usually naturally begins to use English as her or his working language. Such a scholarly form of life and its practices of language-use must also be taken seriously for the very same Wittgensteinian reasons that may lead one to pay attention to philosophers' native languages as a (transcendental) source or ground of their philosophical reflections.

A caveat is in order before we proceed to more specific discussions of Finnish analytic philosophers. Even though Finnish philosophy represents only a small fragment of the large international movement known as analytic philosophy, it is still a relatively comprehensive tradition which I cannot deal with in any detail at all within the scope of one paper. Hence, no one should read the present contribution as an attempt to define "Finnish analytic philosophy" or even to sketch its main historical stages. Much more detailed expositions of various aspects of this tradition can be found in a new collection of essays on the topic (Niiniluoto and Haaparanta 2001, forthcoming; see also Niiniluoto 2000).

2. Rationality and "public use of reason"

The remarks on academic and popular philosophy made above lead us to one of the main themes I wish to take up in this paper: the relation between academic analytic philosophy and the more general "public use of reason", as seen in the geographical context I am working within here. We might say that analytic philosophy, very generally speaking, continues the Enlightenment (modernist) tradition with its insistence on reason, knowledge, truth, and intellectual responsibility. The examination of the norms of rationality rooted in the laws of our logic (and manifested in ordinary language, too) can be regarded as a continuation of the Enlightenment project. Now, this project is also characterized by its emphasis on the public use of reason. The scientific world-view and its rational standards ought to be taken out of the researchers' laboratories and solitary chambers into the real world and be made available to all people. Even though most of the work done in the tradition of analytic philosophy has been rather technical in nature (in Finland and elsewhere), the tradition seems to be committed to this progressive spirit. Analytic philosophers have usually viewed science as a good thing, thinking that its canons of rationality, if not their technical logical formulations, should be exposed in public discussions. For this reason, several philosophers have felt the need to "popularize" their research. In the case of Finland, the distinction between a "scientifically" philosophical paper or book and a popular book or article is perhaps sharper than in English-speaking countries, because the former are usually written in English and the latter in Finnish (or sometimes in Swedish, which is also an official language in Finland, spoken as a mother tongue by a five percent minority). The more technical and formally logical one's academic work is, the greater is, of course, the distance between that work and its popular presentations.

In Finland, several analytic philosophers have, in addition to their academic work, also been active public advocates of rationality. For example, Eino Kaila and Ilkka Niiniluoto have been widely known not only for their defense of science but also for their critiques of religion. Let us first turn to Kaila's work for a moment and then return to more recent issues through a brief discussion of Niiniluoto.

As I already mentioned, Kaila, who was Professor of Philosophy at the University of Turku in the 1920s and Professor of Theoretical Philosophy at the University of Helsinki from 1930 to 1948 (and after that, until his death in 1958, a highly respected member of the Academy of Finland), had contacts with the Vienna Circle during its flourishing period (cf. here the essays collected in Niiniluoto *et al.* 1992). He never endorsed the Circle's most restricted form of empiricism (i.e., the verificationist theory of meaning), and he used to call his view "logical empiricism" instead of "logical positivism". Indeed, he was in many respects closer to scientific realism than to positivism, insisting that the problem of reality (rather than the problem of the meaningfulness of linguistic expressions) was his main concern. Yet, his main philosophical effort was to introduce *scientific philosophy* with its logical and empirical spirit in his home country – and it is, of course, this effort that was continued by his influential followers, especially von Wright and Hintikka (whose contributions to logic have been much more original and important than Kaila's, who was more an epistemologist and philosopher of science than a logician).

While Kaila did write, in German, several academically interesting contributions to logical-empiricist thought (cf. the English translations of his central papers in Kaila

1979), his most lasting achievements were presumably his public uses of reason on the pages of his Finnish monographs, such as *Persoonallisuus* ("The Human Personality", 1934), a work popularizing modern scientific psychology, *Inhimillinen tieto* ("Human Knowledge", 1939), a book on the basic ideas of logical empiricism, and especially his dialogical *weltanschaulich* bestseller, *Syvähenkinen elämä* ("Deep-Mental Life", 1943; first published in Swedish as *Tankens oro*, "The Restlessness of Thought"; an enlargened Finnish edition was published posthumously in 1986). Through these and some other books (not to forget his lectures at the University, which many pupils have regarded as legendary), Kaila's influence on Finnish thought and culture especially from the 1930s to the 1950s was enormous (for selections of his writings in Finnish, see Kaila 1990-92). He practically speaking educated a whole generation of scientists and philosophers, and among his pupils were von Wright, Ketonen, and Stenius, all of whom later became professors of philosophy at the University of Helsinki. (Even Hintikka began his studies with Kaila, although he was primarily von Wright's pupil.) In a word, Kaila and his followers substituted analytic philosophy and its scientific-mindedness for the older Finnish tradition which was oriented to German idealism and Hegelianism.

Now, should we say that Kaila's thoughts directed to a large educated audience did not properly represent his (analytic) philosophy? I do not think so. His writings in Finnish (and partly in Swedish) were an integral part of his philosophical career, in some cases (perhaps in most cases) even more important, philosophically, than the "scientific" publications he produced in German. Nor were they "mere popularizations", although in some cases they were attempts to explain, avoiding technicalities, the basic ideas he tried to develop in his more "scientific" works. Kaila is a splendid, and rare, example of someone truly able to combine academic research with a publically interesting and culturally influential manner of writing.

It is, therefore, hardly surprising that he has had followers in Finland. To mention only one example, Ilkka Niiniluoto (one of my own teachers in philosophy) is internationally (relatively) well-known for his research on the Popperian notion of *truthlikeness* (or verisimilitude) – although it should be noted that his definition of this notion is significantly different from Popper's. In Finland he is, however, much better known for his continuous public struggle in favor of rationality and science, as opposed to various kinds of irrationalism, including (in his view) religious belief. Again, there is a clear connection between the two aspects of a philosopher's work. The concept of truthlikeness is needed, according to Niiniluoto, in order to defend a critical and fallibilistic version of *scientific realism*, which admits that our theories are typically false but can be more or less truthlike descriptions of the structure of a mind- and theory-independent world (see Niiniluoto 1999). If we are willing to respect the scientific pursuit of objective truth, we should embrace "critical scientific realism" with its commitment to the ideals of rationality, progress, and increasing verisimilitude, thereby rejecting inherently irrational and non-progressive belief systems like religions. This, roughly, is the way in which Niiniluoto standardly argues. His technical work in logic and philosophy of science focuses on a number of very detailed questions concerning truthlikeness and some other notions (e.g., the Tarskian definition of truth, probability, induction, the reference of theoretical terms, etc.), and his more popular presentations rely on the technical work by taking for granted a scientific realism on the basis of which it is possible to criticize pseudo-scientific and religious ways of thinking.

I do have my reservations: I believe that Niiniluoto's realism is problematic in many ways and that his (as well as some other Finnish analytic philosophers') atheism is rather superficial if compared to the recent debates in the philosophy of religion (see Pihlström 1996, 1998). There is no need to dwell on these issues here, though. What we should observe is, rather, the intimate relation between academic analytic philosophy and more public, more popular, philosophizing. It is no less important for Niiniluoto's career as a philosopher (or even as an analytic philosopher committed to the ideals of truth and rationality, continuing the project of the Enlightenment) to use his rational capacities in front of a large audience, responsibly trying to educate his cultural peers as well as the yonger generation, than to continue his technical work devoted to the minutiae of truthlikeness and related difficult notions. It is simply for geographical reasons – because he lives where he lives – that he has to do the former in Finnish and the latter in English. For an English-speaking philosopher, there is no such need to split her or his activities into two separate, linguistically identifiable practices. Even if such a split is necessary in non-English-speaking countries like Finland, the issues themselves recognize no sharp divisions. For instance, Niiniluoto's defense of a realism that acknowledges the truth-conduciveness of scientific rationality with the help of the notion of truthlikeness is immediately related to his concern with the irrationality of religious and other non-scientific modes of thinking, which he wishes to give up.

We may briefly mention some other examples of the public use of philosophical reason among Finnish philosophers close to the analytic tradition. In addition to Kaila, the highest public respect and admiration – though also occasional critique – have probably been directed to von Wright's attempts, beginning already in the late 1940s, to diagnose the problematic cultural, political and ecological situation of our Western civilization. Most of his "cultural critique" has been published in Swedish and Finnish, but some of the relevant writings are available in English, too (see von Wright 1993). While continuing the Enlightenment tradition, insisting on the ineliminable importance of rationality in human life (including the public use of reason), von Wright has arrived at a pessimistic attitude toward the ability of our modern culture to solve its problems. He does not believe in reason as a hope for the human race any longer. In a way, as he has admitted himself, he has come to endorse some postmodernists' claims about grand narratives being dead – albeit pessimistically, without sharing the euphoria that postmodernists themselves have often attached to such slogans. von Wright has not been eager to say anything definite about the relation between his analytic philosophical work and his more general cultural concerns, but it has been suggested that an important bridge between the two aspects of his philosophical career could be found in the notion of humanism (cf., e.g., some recent essays in Egidi 1999). Both elements of von Wright's philosophical work have over the decades been focused on human action and its intelligibility.

The requirement that philosophers should use their rational capacities of critical evaluation of various views and arguments in public debates has perhaps nowhere been more explicitly expressed than in discussions concerning "applied ethics". Some Finnish philosophers – most notably, Timo Airaksinen, Heta Gylling (former Häyry), and Matti Häyry – have in the 1980s and 1990s been active in this regard, too. Applied ethicists believe that philosophers can intervene in people's and societies' everyday moral dilemmas

(e.g., problems concerning health care practices or environmental issues) – if not to solve those problems, then at least to conceptually clarify them and to inform the moral agents themselves what their basic options and argumentative strategies look like. Applied ethicists can also try to determine what kind of practical policies of action follow from certain moral theories. There is, however, some reason to suspect the desirability of this form of the Enlightenment ideal of public use of reason. Many Wittgensteinian thinkers – in Finland, Lars Hertzberg in particular – have heavily criticized the presuppositions of applied ethics, arguing that philosophers possess no such special competence which would entitle them to solve or clarify the highly personal moral problems people face. G.H. von Wright himself (influenced by Wittgenstein, as is well known) has also expressed some suspicion to the kind of intellectualization of moral problems that one can find in contemporary work in applied ethics. More generally, there may be difficulties inherent in the very notion of "applying" philosophy (cf. Pihlström 1999). It is not unproblematic to assume that one can first possess a "pure" philosophical (e.g., ethical) theory and then apply it in the practical circumstances of human life. Philosophical conceptions and theories themselves may always already be informed by that life and its problems. This Wittgensteinian theme has perhaps inadequately been emphasized in the analytic tradition. It is certainly closely connected with an idea that was already mentioned above, namely, the view that one's native language and the form(s) of life it is naturally related to may be seen as transcendental conditions for the possibility of (academic) philosophical problems and theories.

In other words, the relation between philosophy and human life is, arguably, much more complex than those who speak about applying philosophy (or, more specifically, about applying ethics) usually admit. It may be interesting to note that one of the most recent "practical applications" in Finnish philosophy is the phenomenon known as "philosophical counseling". The first commercial "philosophical practice" was opened in Finland by Arto Tukiainen in 1999. Since the practitioners of philosophical counseling are usually not very closely associated with the analytic tradition (though Tukiainen himself wrote his dissertation on Wittgenstein), we may perhaps ignore this movement here; on the other hand, I do not want to say anything against it, either, without a more detailed study. In any event, this new philosophical fashion seems to be a result of the growing need to make philosophy more relevant in people's lives and in the society at large. Analytic philosophy will inevitably be affected, directly or indirectly, by this development.

3. Pragmatism, the analytic tradition, and "post-analytic" thought

Our brief discussion of Finnish analytic philosophy would be far from complete if we did not recognize that the conception of rationality favored by most analytic philosophers and the analytic ideal of philosophy itself – however broadly defined in order to include more popular works like Kaila's, von Wright's and Niiniluoto's books and lectures in Finnish and Swedish – have been strongly questioned by what may be labeled the "post-analytic" turn of our analytic tradition. It is this questioning that we should now, finally, turn to for a moment (see also Pihlström 1998, especially ch. 9). I must leave aside the various *anti*-analytic movements in Finnish philosophy that have taken place since the early reception of the analytic tradition by Kaila and others. Those movements have, of course, been directed against the dominance of analytic philosophy, but they are now, on my estimation, largely *passé*, as the analytic school itself is becoming more and more fragmented *from within* (hence the term "*post*-analytic"). Since Sven Krohn's (1949) aggressive attack on logical empirism (including Kaila's views), Finnish philosophy has had its anti-analytic warriors resisting the power of the analytic establishment (cf. here especially Salmela 1998; see also Niiniluoto 2000). (Krohn later became Professor of Theoretical Philosophy at the University of Turku.) Yet, nowadays most people seem to realize that it is more fruitful to build bridges between rival traditions, such as analytic philosophy and (say) phenomenology.

In the international scene, one particularly promising bridge-building tradition since the early years of the twentieth century has been *pragmatism*. Leading American post-analytic thinkers like Richard Rorty and Hilary Putnam (also widely read in Finland) have, at least since the early 1980s, drawn important insights from the tradition of pragmatism and tried to develop quite different versions of neo-pragmatist thought. Now, it seems to me that certain features of Finnish post-analytic philosophy may, in a parallel fashion, be illuminated through a case study of the influences of pragmatism in Finland. In particular, Esa Saarinen's "media philosophy", developed in the 1980s and 1990s, can be regarded as an extreme form of pragmatism giving up traditionally philosophical (normative) considerations of rationality. It can, I think, be shown that pragmatism had a more positive influence on earlier Finnish philosophy – especially on the "pre-analytic" thought of the young Kaila and some of his contemporaries. Pragmatistic ideas are not entirely invisible in the mature (analytic) Kaila, either. (Cf. here Pihlström 2001.)

Saarinen began his philosophical career in the 1970s as Hintikka's pupil, publishing papers on philosophy of language in general and game-theoretical semantics in particular. In the 1980s, he became interested in non-analytic traditions, particularly existentialism, which was not well known in Finland until Saarinen's monograph on Sartre appeared in 1983. During that decade, he also gradually turned into a media celebrity, as well known among the general (non-philosophically inclined) public as any rock star or movie actor, and that aspect of his career became the dominant one in the 1990s (a decade in which one may say he abandoned analytic philosophy altogether). After having been a regular commentator of almost any cultural and social phenomenon one might imagine in the Finnish newspapers and the electric media (e.g., sex, rock music, fashion, fast food, political elections, business management), Saarinen has now become a private lecturer and management consult whose services are purchased even by major Finnish companies, especially Nokia. His insistence on media philosophy – on the philosopher's active engagement in public debates in the electric media, comparable (in his view) to Socrates's discussions at the *agora* of Athens (cf. Taylor and Saarinen 1994) – has thus been accompanied by his conception of philosophy as a commercial "service industry" (Saarinen 1998). Since Saarinen's customers are large companies rather than individuals needing "philosophical therapy", his work is not usually classified as philosophical counseling. Nor are applied ethicists happy about his rather liberal use of the term "applied philosophy" as a description of what he is doing.

Saarinen's new (clearly non-analytic) work is post-analytic in the sense that it grows out of his dissatisfaction with the irrelevance of the technicalities of analytic philosophy,

to which he was himself busily contributing during his early years as a professional analytic philosopher in the 1970s. Thus, his more recent conception of philosophy can be associated with pragmatism (that is, if we decide to regard what he is doing as philosophy at all). His main thesis seems to be that philosophical ideas ought to be put to work in people's lives: philosophy should be relevant to people's problems and make people (not only academic philosophers) happier. There are, obviously, several problems with this radically pragmatist way of thinking. There seems to be no critical control regarding the sort of "work" that the (so-called) philosophical views of life Saarinen advances can do. His message appears to be the rather simple one celebrating happiness, active life, and "flourishing", with no genuinely philosophical argument or analysis problematizing these notions. Indeed, one may ask whether there are any criteria defining "philosophy" any longer, if anything a philosopher says in the media or in front of an audience consisting of business managers (or whoever) can be considered philosophical.

We cannot here analyze Saarinen's ultra-pragmatist work and its problematic metaphilosophical assumptions any closer (cf. Pihlström 1998, 2001). Suffice it to say that while I find pragmatism in general an interesting and highly promising philosophical tradition, with several significant connections to analytic philosophy (early and late), I am tempted to judge Saarinen's media-philosophical ideas as far too radically pragmatistic. What Saarinen offers us is a degraded, radically relativistic transformation of the (respectable) pragmatic insistence on the relevance of philosophy in human life.

Saarinen has not explicitly connected his work with the tradition of pragmatism, but some earlier Finnish philosophers made more interesting philosophical use of that tradition already before the rise of the analytic tradition as the dominant paradigm of Finnish philosophy (see Pihlström 2001). Kaila, in particular, was interested in William James's thought in the 1910s (i.e., before he became a logical empiricist and analytic philosopher), and even reviewed several Finnish translations of James's works published at that time. Later, James's influence can be seen in some of Kaila's popular works, particularly *Syvähenkinen elämä*. Surely, Kaila cannot be considered a pragmatist, but he took seriously the Jamesian pursuit of formulating such a philosophical *Weltanschauung* that may enable one to be "at home" in the world, i.e., a philosophy relevant to one's life. The project may not be an unproblematic one, but it is intellectually more admirable – and certainly more clearly a part of our Western tradition of philosophy – than Saarinen's rather vulgar pragmatism.

The key difference between the two lies, perhaps, in the role played by the notions of rationality and irrationality. In intellectually responsible forms of pragmatism and post-analytic thought, the significance of these notions is not obscured, even if their traditional analytic characterizations are (sometimes heavily) criticized. Post-analytic philosophy (including pragmatism) is, in its most interesting versions, an attempt to investigate the limits of rationality as it has been conceived in the logical-analytic school.

4. Conclusion

It is clear, in my view, that academic work in analytic philosophy (or in some other tradition of academic philosophy, such as phenomenology) *may* lead to interesting and intellectually responsible (though undoubtedly controversial) public uses of reason. Such attempts to use reason philosophically in front of a larger audience than the academic one should, however, maintain some connection to serious philosophical engagements with the notions of rationality and irrationality. Here even quite traditional analytic philosophy can turn out to be fruitful from the point of view of the Enlightenment ideal of rationality it correctly emphasizes. What should be avoided is the false choice between *either* dry academic work that has no connections whatsoever with (what John Dewey called) the "problems of men" *or* irresponsible public speculation and sophistry that lacks intellectual substance. The increasingly technical and professionalized analytic philosophy may have committed the former sin, whereas some radical post-analytic thinkers like Saarinen are perhaps, unfortunately, close to having committed the latter.

I wish to conclude by suggesting that the emergence of neopragmatist and post-analytic movements – in Finland and elsewhere – ought to be critically examined rather than entirely rejected, even if (or especially if) we want to continue rather than bury analytic philosophy. They may offer valuable material for an investigation of the notions of reason and rationality. In Finland, at least, analytic philosophy has gradually been forced to adopt a more open attitude to non-analytic ways of philosophizing. There is no unified school-like institution of analytic philosophy any longer. Moreover, analytic philosophers themselves nowadays recognize that they work within a historically developing tradition instead of occupying some imagined ahistorical standpoint. While Kaila's and his followers' "scientific philosophy" gradually took the place of the more historically oriented German school which had earlier been strong in Finland, the analytic and post-analytic philosophers of the 21st century cannot fail to take the history of philosophy seriously – including the history of their own tradition.

Yet, I do not want to present too happy a picture of the future prospects of Finnish analytic and post-analytic philosophy. In our country, as in most other countries dominated by the analytic paradigm in the 20th century, the most serious danger to be faced in the future is unreflective, unphilosophical scientism. Even though the specific doctrines of the Vienna Circle have been abandoned long ago, the spirit of the Circle is still strong among analytic philosophers, most of whom even in these post-analytic days believe in the ultimate triumph of a reductionist natural-scientific picture of the world. The scientific reductionism and physicalism of mainstream analytic philosophy is, in my view as well as in many others', an unhealthy tendency that should be criticized from within the tradition of analytic philosophy itself (as well as from the point of view of pragmatism, for instance). It may be the case that the increasingly strong habit among Finnish philosophers of adopting their favorite philosophical outlooks from the United States is not entirely unrelated to the scientistic developments typical of recent analytic thought. On the other hand, the various post-analytic critiques of scientistic currents of thought often have an American origin, too. Whether or not one believes in a unified naturalistic, scientifically reductionist conception of reality, most of the books one reads, either approvingly or disapprovingly, are written by American philosophers. Finnish philosophy is nowadays so much dependent on American (analytic and post-analytic) thought that it is sometimes rather odd even to speak about a national tradition any more. Even so, the development of analytic philosophy in Finland constitutes a historical phenomenon that is interesting from the point of view of national Finnish culture, too. Figures like Kaila, von Wright,

and Niiniluoto have been extremely influential in Finnish cultural life during the last century.

In any event, what neither analytic nor post-analytic thinkers, neither pragmatists nor their opponents, neither reductionists nor their critics, should forget in the contemporary philosophical situation troubled by both scientistic and irrational, anti-scientific threats is the task of explicating philosophically (as well as historically) the concept of rationality. This is an equally compelling task (for Finnish and non-Finnish philosophers alike), whether or not one believes in the orthodox analytic idea of the norms of reason being grounded in our basic logic itself.

References

Egidi, R. (ed.) 1999 *In Search of a New Humanism: The Philosophy of Georg Henrik von Wright*, Dordrecht: Kluwer.

Kaila, E. 1979 *Reality and Experience*, Dordrecht: D. Reidel.

Kaila, E. 1990–92 *Valitut teokset* (Selected Works, in Finnish), vols. 1–2, ed. I. Niiniluoto, Helsinki: Otava.

Krohn, S. 1949 *Der logische Empirismus: Eine kritische Untersuchung, Erster Teil*, Annales Universitatis Turkuensis B 31, Turku: Turun yliopisto.

Niiniluoto, I. 1999 *Critical Scientific Realism*, Oxford and New York: Oxford University Press.

Niiniluoto, I. 2000 "From Logic to Love: The Finnish Tradition in Philosophy", manuscript, updated (earlier version published in the *Proceedings of the Estonian Academy of Sciences*, 1993).

Niiniluoto, I. and Haaparanta, L. (eds.) 2001 (forthcoming), *Analytic Philosophy in Finland*, Poznan Studies in the Philosophy of the Sciences and the Humanities, Amsterdam and Atlanta: Rodopi.

Niiniluoto, I., Sintonen, M. and von Wright, G.H. (eds.) 1992 *Eino Kaila and Logical Empiricism*, Acta Philosophica Fennica 52, Helsinki: The Philosophical Society of Finland.

Pihlström, S. 1996 *Structuring the World: The Issue of Realism and the Nature of Ontological Problems in Classical and Contemporary Pragmatism* (Acta Philosophica Fennica 59), Helsinki: The Philosophical Society of Finland.

Pihlström, S. 1998 *Pragmatism and Philosophical Anthropology: Understanding Our Human Life in a Human World*, New York: Peter Lang.

Pihlström, S. 1999 "Applied Philosophy: Problems and Applications", *International Journal of Applied Philosophy*, 13, 121–133.

Pihlström, S. 2001 "Pragmatistic Influences in Twentieth Century Finnish Philosophy: From Pre-Analytic to Post-Analytic Thought", forthcoming in Niiniluoto and Haaparanta 2001.

Saarinen, E. 1998 "Philosophy as a Service Industry", paper presented at the 20th World Congress of Philosophy in Boston, August 1998.

Salmela, M. 1998 *Suomalaisen kulttuurifilosofian vuosisata* (in Finnish; English abstract: Cultural Philosophy in Finland in the twentieth Century), Helsinki: Otava.

Taylor, M.C. and Saarinen, E. 1994 *Imagologies: Media Philosophy*, London and New York: Routledge.

von Wright, G.H. 1993 *The Tree of Knowledge and Other Essays*, Leiden: E.J. Brill.

Why It's Irrational to Believe in Consistency

GRAHAM PRIEST

1. Introduction

When we describe someone as being inconsistent, we may mean many different things. If someone is fickle, changing their views or emotions from day to day, we call them inconsistent. If someone, such as a judge, fails to treat like cases alike, we call them inconsistent. When someone refuses to accept an obvious logical consequence of something they believe, we call them inconsistent. And of course, when someone believes something of the form $A \wedge \neg A$ we call them inconsistent. In all such cases, calling a person inconsistent would normally be some form of criticism. The first three cases will not concern us here; only the last. The topic will be precisely whether, in this case, such criticism is perforce justified. I shall argue that it is not. Inconsistency is not necessarily rationally vicious: – on the contrary, it may be virtuous.

2. The History of the Belief in Consistency

Let us start with the history of the view that criticism on the ground of inconsistency (in the appropriate sense) is rationally mandated. This is a curious one, not to say a little puzzling. There were certainly Presocratic philosophers who, presumably, thought it was perfectly legitimate to believe contradictions, since they did so. Heraclitus, for example, held that 'We step and do not step into the same rivers; we are and we are not'.[1] Even Plato seems to have been prepared to countenance the possibility that ordinary things might have contradictory properties – though not the forms:

> Even if all things come to partake of both [the form of like and the form of unlike], and by having a share of both are both like and unlike one another, what is there surprising in that? ... when things have a share in both or are shown to have both characteristics, I see nothing strange in that, Zeno, nor yet in a proof that all things are one by having a share in unity and at the same time many by sharing in plurality. But if anyone can prove that what is simple unity itself is many or that plurality itself is one, then shall I begin to be surprised.[2]

Of course, interpreting texts such as these, especially the Presocratics, is a sensitive matter. But at least Aristotle interpreted his Presocratic precursors as endorsing contradictions. For in Book 4 of the *Metaphysics* (especially Chapter 4), he launched a sustained attack on them, claiming that nothing is more certain, better known, or epistemically more fundamental than that a contradiction cannot be true, the Law of Non-Contradiction (LNC). In fact, it is so fundamental that one cannot give a proof of the Law (1005^b9–6^a11). From this, it follows that it is irrational to endorse a contradiction. For it is irrational to believe something that is certainly false.[3]

Now, the claim that the Law is certain is an odd one. If this is so, how come this wasn't obvious to the people Aristotle was attacking? It clearly wasn't. Worse, despite what Aristotle says about the impossibility of proving the Law, he immediately goes on to give seven arguments for it. He calls these elenchic arguments rather than proofs. What this means, and whether all the arguments are elenchic, is not at all clear, but we do not need to go into that. The proofs are no mere intellectual exercise. Their rhetorical content is high. Opponents are likened to vegetables and ridiculed for not practising what they preach (1008^b3–31). These arguments are meant to persuade. No one argues like this for something that is really certain.

Perhaps, then, the Law is certain in the light of Aristotle's arguments? Hardly. The first, and longest, argument is so opaque that it is entirely unclear what it is. I defy any sensible person to read it and claim that it makes its conclusion obvious. The other arguments are even less successful: for their conclusion is patently that it is not the case that *all* contradictions are true (or the even weaker: it is impossible for someone to believe that all contradictions are true). Even if this is true, and were manifestly so, nothing at all follows about the possibility that *some* contradictions are true.[4]

At various times since Aristotle, various thinkers have wittingly endorsed contradictions. This is particularly true of thinkers in the Neoplatonist tradition. (I include Hegel in this number.[5] Plotinus himself, for example, denies that anything positive can be said of the One (*Ennead* V, 6); but also describes it as a simplex, something which is beyond being, the source and generator of all else. The contradiction must have been obvious to him. Cusanus goes even further, saying that (*Of Learned Ignorance* I, 3. Translation from Heron (1954)):

> in no way do they [distinctions] exist in the absolute maximum [*sc.* God]; The absolute maximum ... is all things, and whilst being all, is none of them ...

But it is fair to say that, at least since the middle ages, Aristotle's views concerning contradiction have been high orthodoxy. (This is so obvious, that it is hardly worth documenting.) They are taken for granted so much that, as far as I know, there is no sustained defence of the LNC in Western philosophy other than Aristotle's. Why?

I really don't know. It is certainly not because of the rational persuasiveness of Aristotle's arguments. My conjecture is that the view became entrenched when Aristotle replaced Plato as the dominant philosophical authority in the medieval European universi-

1. Fragment 49a; translation from Robinson (1987), p. 35.
2. *Parm.* 129b, c. Translation from Hamilton and Cairns (1961). The rest of this puzzling dialogue does seem to surprise Socrates in just this way.
3. Though one might take issue with this claim. Maybe it is not irrational to believe something that is certainly false, if it can be shown to be true too. See Priest (1993a).
4. The arguments are discussed in detail in Priest (1998).
5. For the influence of Neoplatonism on Hegel, see Kolakowski (1978), ch. 1.

B. Brogaard, B. Smith (Hg.), *Rationality and Irrationality / Rationalität und Irrationalität*, S. 284–293.

ties, and that the views on contradiction were just taken on as part of that authority. The magisterial position of Aristotle disappeared long ago, of course. In logic it hung on till the 20th century; most of that has been swept out since then, but the views about contradiction have hung on tenaciously.

Before we leave the history of philosophy, and to emphasize the parochial nature of Western Philosophy on this issue, let us consider briefly the situation in Indian Philosophy. The LNC certainly had its defenders here. Logicians of the Nyaya school, for example, adhered to it.[6] But the orthodoxy in this case was exactly opposite from that in the West. The standard view in Indian logic, going back to about the same time as Aristotle, was that on any issue there are always four possibilities to be considered: that a view is true (and true only), that it is false (and false only), that it is both true and false, and that it is neither. This is the *catushkoti* ("four corners"). In other words, the possibility that both a claim and its negation are true, is standard fare. The difference in the Western and Indian traditions could not be more stark.

So far, then, we may conclude that if figures in the history of Western philosophy held that it was irrational to believe a contradiction, this view was itself irrationally held. Though people may have thought it to be so, the LNC is not at all obvious – either in itself, or in the light of arguments that were given. Wise people, as Hume put it, apportion their beliefs according to the evidence. Those who subscribed to the orthodox view, were not, then, wise. They instantiated a pathology of reason.

A caveat: no one is an expert in many areas, and we all have to accept things on the say-so of experts in some areas. If someone believes that it must be irrational to believe contradictions, on the basis of the pronouncements of experts in logic and philosophy, their belief is not irrational. The charge of irrationality is levelled against people who ought to have known better.

3. Is Consistency Rationally Mandatory?

Is the situation essentially any different in contemporary philosophy? Not as far as I can see. The knee-jerk reaction of most modern philosophers will be that no contradiction can be true, since, if it were, everything would be true, which it is patently not – everything follows from a contradiction (*Explosion*). But this conclusion is only as sure as Explosion. This holds in standard modern logic, of course; but it does so only because of the fact that in the semantics of this logic, truth and falsity are taken to be mutually exclusive – something that may fail in a paraconsistent logic, where contradictions do not entail everything. The assumption that truth and falsity are exclusive is simply packed into nearly all presentations of standard logic without comment. In other words, it has the same dogmatic status as the LNC itself. This does not provide a justification.[7]

6. And, as far as I know, for reasons not dissimilar to those of Aristotle. For example, the 11th century Nyaya philosopher, Udayana, sketches an argument for the LNC from the threat of failure of intelligibility of speech, in case contradictory expressions are admitted into the legitimate part of the language. (Thanks to Jay Garfield for this information.) This sounds very much like one interpretation of the most important of Aristotle's arguments.
7. It is a curious fact that Aristotle defends the Law of Excluded Middle in the same book of the *Metaphysics* as he defends the Law of Non-Contradiction, but he appears go back on this in Ch. 9 of *De Interpretatione*. It would seem that he never went back on what he said about the Law of Non-Contradiction.

It might be thought that the exclusivity of truth and falsity holds by definition: falsity just is the lack of truth. It does not. The relevant sense of falsity here is truth of negation; and the claim that $\neg A$ is true when A fails to be true is *not* a definition. It is a substantial theory about the way that negation works. It is denied in modern logic by paraconsistent logicians and intuitionist logicians. How negation behaves is, in fact, a highly contentious issue historically. At least until the middle ages, a common view of negation, with which Aristotle had some sympathies, was that $\neg A$ simply cancels out A. So contradictions do not entail everything: they entail nothing.[8] The matter is not, therefore, one that can simply be settled by definition.

What other relevant arguments are there? It is hard enough to produce arguments for the LNC itself, but what we need in this context is not just an argument for the LNC, but one of a very strong kind. Consider, as an analogy, the arguments given by the early heliocentric astronomers for the centrality of the sun. These had force, but there were countervailing arguments; for example, the motion of the earth seemed to fly in the face of the accepted dynamics. Given this, it was not irrational for someone to accept the geocentric view. It would have been, had the arguments for heliocentrism rationally mandated their conclusion, something that later arguments were to do. In the same way, and returning the issue of inconsistency, what would need to be found in this case is not simply an argument for the LNC, but one that makes its conclusion rationally mandatory. This sets the bar very high, and I know of no argument, or raft of arguments, that comes even close to clearing it.

Let me finish this section with a couple of comments on a lesser form of inconsistent belief. It would appear quite possible for someone to believe A and $\neg A$ without believing their conjunction. Arguments for the irrationality of this sort of inconsistency are even harder to find. For in such situations, there is nothing that is manifestly false (as $A \wedge \neg A$ is – supposedly), which the person believes. Perhaps the best one can do to establish the irrationality of this sort of situation is to reduce it to the conjoined case by an additional argument to the effect that a rational person should believe all the (obvious?) logical consequences of their beliefs. However, even this additional argument is hard to sustain. If a high enough probability is a sufficient condition for believing something, then, as the "lottery paradox" demonstrates, it may be rational to believe things without believing their conjunction. At any rate, if it is not necessarily irrational to believe $A \wedge \neg A$, it is certainly not necessarily irrational to believe A and $\neg A$.

4. Cognitive Virtues

Those who take consistency to be a *sine qua non* of rationality may well feel discomforted by all this. If the inconsistency of a view does not show it to be irrational, what does?! In fact, even if consistency were a constraint on rationality, it would be a relatively weak one. It is possible to massage many rationally bizarre views into consistent ones. For ex-

8. For details, see Priest (1999c).

ample, the view that the earth is flat can be held quite consistently, by invoking suitable auxiliary hypotheses about the behaviour of light, a conspiracy of the world's media, etc. It is irrational for all that. So how does rationality work? Anyone who expects a simple algorithm to determine whether a belief is rational or not is bound to be disappointed. This is a lesson of post-positivist philosophy of science, if, indeed, it was not already to be learned from Aristotle's account of *phronesis*.

Given a theory (in any area of human cognition – science, philosophy, logic, or whatever), there are many cognitive virtues, and, correlatively, vices, that it may have. Perhaps the most important of these is the adequacy of the theory to the data which it was proposed to handle. Does it really account for these? Any other criteria are contentious, but familiar candidates include, for example:

Simplicity: Is the theory clean and elegant, or it is contrived and kludgey?

Unity: Does the theory have to invoke numerous *ad hoc* hypotheses, coming in from left field?

Parsimony: Does the theory multiply entities beyond necessity?

It is clear that all these criteria may come by degrees. And as even a quick perusal of some intellectual history demonstrates, they may pull in different directions. The early Copernican theory was simpler than the Ptolemaic theory, but it could deal with the dynamic problems of the motion of the earth only in *ad hoc* ways, at least until the invention of a new dynamics. By contrast, the Ptolemaic theory, though more complex, was not *ad hoc* in this way.[9]

When is one theory rationally preferable to another? When it is sufficiently better than its rivals on sufficiently many of these criteria. That is, of course, vague. It can be tightened up in various ways.[10] But in the end, I think, it is essentially so. Indeed, it is precisely this vagueness that allows for rational people to disagree. For legitimate disagreement is precisely a feature of the borderland between R and $\neg R$ for a vague predicate, R. People may legitimately disagree, for example, over whether the application of 'child' to a certain 14 year old is correct. This does not mean that there are no determinate facts concerning rationality and similar vague notions. Someone who calls any 60 year old a child is clearly mistaken. And one theory *can* be manifestly superior to its rivals, all things considered.

Is consistency a cognitive virtue? Consistency may certainly be required by other virtues. For example, if the theory in question is an empirical one, adequacy to the empirical data is certainly a virtue. But at least for the most part, such data are consistent: contradictions are rarely perceived in the empirical world, and where they are, they are illusions. Hence, adequacy to the data requires consistency of empirical content.[11] But is consistency *per se* a cognitive virtue? I am not sure of the answer to this. The issue cannot be divorced from the question of what makes something a cognitive virtue. This is an exceptionally hard question. Why, for example, is simplicity a virtue? I don't know the answer to that either. If there are reasons, perhaps of a transcendental kind, for supposing, quite generally, that the world (all that is the case) has a low degree of inconsistency, then consistency is a cognitive virtue. If not (and I know of no such reasons), then perhaps not.

Whether or not consistency is a cognitive virtue is, however, not centrally important for the present matter. Whether or not it is, it is simply one virtue amongst many. It is clear, then, how an inconsistent theory may be rationally acceptable: it scores more highly than its rivals on the (other) cognitive virtues. Conversely, it is clear how an inconsistent theory may be rationally rejectable. It may simply be trumped by another theory scoring more highly.

Since this latter point is something of a stumbling block in discussions of paraconsistency, let me quickly illustrate it. Let us suppose that you, an atheist, argue against the existence of (a Christian) god on the basis of the existence of suffering. God is omniscient, omnipotent, perfectly good, etc. Take some event which we know to have occurred, and which caused much gratuitous suffering, for example the torture and murder of an innocent child. The properties of God entail that had such a being existed, this event would not have occurred. But it did; hence there is no God.[12] Now consider a person who is prepared to accept a contradiction in this context: God exists and prevented the event; hence it did not occur; but it occurred too. This move is certainly one that is out there in logical space. Is it a rational one? Not really; that the event both did and did not occur is an empirical contradiction, and as we have already seen, such contradictions are not acceptable.[13]

5. The Rationality of Inconsistency

So far, I have argued that the fact that someone's beliefs are inconsistent is not necessarily a ground for rational criticism – and more, that to hold that it must be, is itself irrational. In the rest of this paper, I want to argue for something stronger: that it is *irrational* to be consistent. I do not mean that it is irrational to be consistent about everything: just that there are some topics about which it is irrational to be consistent. Nor do I mean that it always has been, and always will be, irrational to be consistent about these things: simply that in the present state of our knowledge, the rational belief is an inconsistent one.[14]

The topic I have in mind here is that of truth. It seems to me that anyone weighing up the state of play concerning this notion ought rationally to be inconsistent. To make the case for this in full would require much more time than I have here. What I will do is make

9. See Chalmers (1976), 6.5.
10. See Priest (2000b).
11. This view is defended at greater length in Priest (1999b).
12. Of course, there are many things one might dispute about such an argument: there was no such event; God had good reasons for letting it happen, etc. These moves are not on the agenda here.
13. One could take on this claim as well, of course, but good luck to someone who does! I see no way of defending the view that the event both did and did not occur without invoking countless *ad hoc* hypotheses, and turning the situation into one of the flat-earth kind.
14. If we are dealing with the weaker form of inconsistency, where a person believes A and $\neg A$ without believing their conjunction, then arguments of a quite different kind can also be employed; for example, the "paradox of the preface". See Priest (1987), p. 124.

the basic case, and consider a few objections. The rest will have to be left to your own cogitations.

First, it is pretty universally agreed that an overwhelmingly natural principle concerning truth is the *T*-schema: for every proposition, *A*, <*A*> is true iff *A* (where angle brackets are some name-forming device). The schema is as ancient as Aristotle, and as modern as deflationist accounts of truth. The most celebrated truth-theorist of the 20th century, Alfred Tarski, even called it a criterion of adequacy on any account of truth. Let us call an account of truth that endorses the *T*-schema a naive account.

The problem with endorsing a naive account is, of course, that, in conjunction with a mechanism of self-reference and a few simple logical principles, it leads to inconsistency in the shape of paradoxes such as the Liar. Such is the intuitive force of the *T*-schema that, I think it fair to say, if it were not for this fact, there would be no dispute concerning the claim that a naive account is correct. If it is irrational to reject a view merely because it is inconsistent, a naive view of truth would seem to be rationally obligatory.

This is a bit too fast, however. The rational view on any matter is the best of all the competing views on offer. What competitors are there presently to a naive view? As most people in the audience will hardly need to be told, there is a plethora. 20th-century logic provided us with accounts by Kripke, Gupta and Herzberger, Barwise and Etchemendy, McGee, and Tarski (notwithstanding his views about the *T*-schema), to name some of the major ones. All non-naive accounts start with a major strike against them. The mere fact that they do not endorse the *T*-schema means that they fail the single most important cognitive virtue: adequacy to the data. At best, they can account for only part of this; some way must be found of writing off the other part. In virtue of this they had better score high on many of the other criteria. Do they?

At this point, we ought really to engage in a detailed analysis of the various accounts, since it is wrong to suppose that they are all of a kind. But in this context some broad brush-strokes will have to suffice.[15] For a start, does any of them score high on the virtue of simplicity? Certainly not in comparison with a naive account. All involve hierarchical constructions, of various degrees of complexity, going into the transfinite. Most importantly, do these accounts themselves avoid being inconsistent? Not really; for all of them, the machinery deployed allows one to construct Liar-type arguments (extended paradoxes) ending in contradiction. We may, in each case, try to avoid these new contradictions; but the moves are, for the most part, *ad hoc*. Worse: it is clear that we are still faced with the essentially the same problem with which we started: the ability of semantically closed language to generate paradoxes. This therefore succeeds in manifesting another cognitive vice. Making this move does not solve the problem: it merely relocates it. Suppose, by analogy, that I want to explain why there is a physical cosmos: I find it difficult to see how there could be something without a cause. To solve the problem, I postulate the existence of a creator-god. It is clear that, though, in a sense, this solves the original problem, in a more important sense, it merely relocates it. If this is all there is to the move, I should be equally puzzled by the existence of a god. In a similar way, if the problem posed by the Liar and its kind is to explain how our semantic notions work in a consistent way, a solution that produces semantic notions that are subject to exactly the same problems, gets us nowhere.[16]

15. For an analysis of Tarski, Kripke, Gupta and Herzberger, see Priest (1987), ch. 2. For Barwise and Etchemendy, see Priest (1993b); for McGee, see Priest (1994).

6. Objections

Now for the objections. All of these have been made – and answered – elsewhere, so I may be brief.

Objection 1: The comparison between a naive account and consistent accounts is unfair. For one needs different underlying logics for the two accounts, a paraconsistent logic for the naive account and classical logic for the consistent account. One needs to evaluate the package deal (truth plus logic) in each case; and this changes the evaluation, since classical logic is much simpler than any paraconsistent logic, making the consistent package over-all simpler.

Reply: The point about the package deal is quite correct. It is also true that classical logic is simpler than all paraconsistent logics – but it is not *that* much simpler in many cases. For classical logic must be taken to include modal logic. We reason with modal sentences all the time, and we certainly need a theory of truth that applies to these. Now the simplest relevant logic is not much more complex than modal logic.[17] Its semantics diverges from those of *S5* in two ways. First, the classical assumption that truth and falsity (at a world) are exclusive and exhaustive is dropped. Technically, this is a very minor change. Next, the class of worlds is extended to include not just possible worlds, but impossible worlds too (and conditionals are given suitable truth conditions at these). But again, impossible worlds are things we need to countenance anyway: they are needed, for example, to provide a suitable semantics for counterfactuals with logically false antecedents. Or, to put it another way, if we give a worlds-account of conditionals, and stick with the classical notion of world, our account will produce most implausible results – such as the truth of 'If you were to square the circle, I would give you my life's savings'. (Compare: if you were to square the circle you would become a famous mathematician.) Extra complexities will then have to be invoked to handle these results. It is not clear, thus, that the move to a relevant logic does increase complexity, all things considered. If it does, the move is not of a kind that changes the over-all simplicity assessment wildly.

Objection 2: Even if a certain amount of inconsistency is acceptable, too much is not. Thus, if inconsistencies spread into the empirical consequences of the theory of truth, the theory is not rationally acceptable. The consequences of a naive theory do spread into this area, due to Curry paradoxes. Specifically, given the inference of contraction $(A \to (A \to B) \vdash A \to B)$, everything follows using the *T*-schema.

Reply: The point about the unacceptability of wide-spread contradiction is correct. However, with the appropriate logic, contradictions are appropriately quarantined. Contraction is not valid in the simplest relevant logics. (Nor is its failure *ad hoc*. It is a simple consequence of the nature of impossible worlds.) It can be shown that the inconsistencies in a naive theory of truth do not spread into the empirical realm; indeed, all of the sen-

16. See Priest (1999a), sec. 6.
17. I refer here to the logic *N4* of Priest (2001), ch. 9.

tences that are grounded (in the sense of Kripke) are contradiction-free.[18]

Objection 3: The employment of a non-classical logic means that the paraconsistent position concerning truth suffers from an important cognitive vice. For such a logic is weaker than classical logic. Hence, much classical reasoning must be accepted as invalid. This means that many of the important applications of classical logic must be given up, producing a significant loss of over-all explanatory power.

Reply: It is true that many inferences that are classically acceptable must be acknowledged as deductively invalid. This does not occasion an explanatory loss, however. For the only situations about which it makes sense to reason classically are consistent ones; and even paraconsistent logicians may employ classical logic in consistent situations (just as intuitionists may employ classical logic when reasoning about finite situations): classical logic is just a special case. The whole idea can be made formally rigorous and precise by the construction of an appropriate non-monotonic logic. The inferences in question are quite acceptable non-deductive inferences.[19]

Objection 4: The naive account of truth is just as susceptible to extended paradoxes as consistent accounts. For we may define an operator, # (whether or not one calls this negation) by the conditions of classical negation, and then use the *T*-schema to infer something of the form $A \wedge \#A$, giving rise to triviality. In the end, then, a naive account is no more acceptable than a consistent account.

Reply: No logical theory of any kind can allow unbridled licence in postulating connectives satisfying arbitrary conditions. The point was forcibly brought home by Arthur Prior in his discussion of *tonk*.[20] From the point of view of the naive theory, any attempt to define a connective obeying the laws of classical negation will either produce a connective that does not satisfy these laws or will not succeed in specifying a meaningful connective at all. Nor is the naive truth theory in the same situation as consistent theories here. For consistent theories entail that the notions employed in formulating extended paradoxes are sensible ones: the notions are part, indeed, of the semantic theories being employed. The naive theory of truth, and the semantics of relevant logics, have no need for #.[21]

7. Conclusion

Whether or not one is persuaded by the details of the example concerning truth, they illustrate at least the possibility that rationality may itself *require* inconsistency. It is a truism to point out that one of the greatest opponents of rationality has always been superstition, often of a religious nature. It is not a truism to note that one of the greatest superstitions in the history of Western thought has been that concerning consistency. Consistency has been taken to be the very corner-stone of rationality. But this view has itself no rational ground: it is simply the legacy of the reverence for Aristotle. An inconsistent view may,

18. See Priest (200a), 8.1–8.3.
19. See, e.g., Priest (200a), 7.6.
20. See Prior (1960).
21. The matters are explained further in Priest (1990) and (1999a).

in fact, be the very embodiment of rationality. It is time to shake off, as Wittgenstein put it, the 'superstitious dread and veneration in face of contradiction'.[22] Enough is enough.

Bibliography

A. Chalmers (1976), *What is This Thing Called Science?*, University of Queensland Press.
E. Hamilton and H. Cairns (1961), *The Collected Dialogues of Plato*, Princeton University Press.
G. Heron (trans.) (1954), *Of Learned Ignorance*, Routledge and Kegan Paul.
L. Kolakowski (1978), *Main Currents of Marxism, Vol. 1: the Founders*, Oxford University Press.
T.M. Robinson (1987), *Heraclitus: Fragments*, University of Toronto Press.
G. Priest (1987), *In Contradiction*, Martinus Nijhoff.
G. Priest (1990), 'Boolean negation and all that', *Journal of Philosophical Logic* 19, 201–15.
G. Priest (1993a), 'Can contradictions be true? II', *Proceedings of the Aristotelian Society, Supplementary Volume* 67, 35–54.
G. Priest (1993b), 'Another disguise of the same fundamental problems: Barwise and Etchemendy on the Liar', *Australasian Journal of Philosophy* 71, 60–9.
G. Priest (1994), Review of V. McGee, *Truth, Vagueness and Paradox*, *Mind* 103, 387–91.
G. Priest (1998), 'To be *and* not to be – that is the answer; on Aristotle on the law of non-contradiction', *Philosophiegeschichte und Logische Analyse* 1, 91–130.
G. Priest (1999a), 'What not? A defence of a dialetheic theory of negation' in D. Gabbay and H. Wansing (eds.), *What is Negation?*, Kluwer Academic Publishers.
G. Priest (1999b), 'Perceiving contradictions', *Australasian Journal of Philosophy* 77, 439–46.
G. Priest (1999c), 'Negation as cancellation, and connexive logic', *Topoi* 81, 1–8.
G. Priest (2001), *Introduction to Non-Classical Logic*, Cambridge University Press, forthcoming.
G. Priest (2000a), 'Paraconsistent logic', in Vol. E2 of D. Gabbay and F. Guenthner (eds.), *Handbook of Philosophical Logic*, 2nd edition, Kluwer Academic Publishers, forthcoming.
G. Priest (2000b), 'Paraconsistent belief revision', *Theoria*, to appear.
A. Prior (1960), 'The runabout inference ticket', *Analysis* 21, 38–9; reprinted in P. Strawson (ed.), *Philosophical Logic*, Oxford University Press, 1967.
L. Wittgenstein (1978), *Remarks on the Foundations of Mathematics*, 3rd edition, Basil Blackwell.

22. Wittgenstein (1978), p. 122.

Rationalität und der Glaube der „religiös Eingeweihten"

EDMUND RUNGGALDIER

1. Einleitung

Aus der umfassenden Problematik der Rationalität religiösen Glaubens greife ich einen Aspekt heraus, der häufig einerseits zu *Verärgerung*, andererseits zu *Frustration* führt. Es ist die Bekundung vieler Menschen, ihr religiöser Glaube sei nur jenen zugänglich und verständlich, die entweder entsprechende Erfahrungen gemacht haben oder auf eine bestimmte Weise leben. Religiöse Menschen bekunden zuweilen auch schlichtweg, sie seien durch besondere Eingebungen zu Einsichten gekommen, die anderen notgedrungen verschlossen blieben. Noch radikaler betonen Mitglieder extremer Glaubensgruppen, daß ihre Glaubensannahmen und -inhalte Nicht-Eingeweihten vorenthalten seien. Ihre Kenntnis und ihr Verständnis setzten eine entsprechende *Einweihung* voraus.

Besonders auffallend ist das Problem in bestimmten Kreisen der diffusen Religiosität esoterischer Provenienz. Da spielt z.B. die sogenannte *Erleuchtung* eine ausschlaggebende Rolle: Wer nicht selber durch regelmäßige Meditation und durch Führung eines Meisters zumindest ansatzweise erfahren hat, worin Erleuchtung besteht, könne über diesen Zustand mit anderen nicht rational sprechen. Bereits die Wurzel des Ausdrucks „Esoterik" (aus dem Griechischen „eis") gibt zu verstehen, daß die esoterischen Lehren nicht jedermann zugänglich sind, sondern nur dem inneren Kreis der Eingeweihten (siehe: Runggaldier 1996).

In der Auseinandersetzung darüber, ob es vernünftig oder intellektuell redlich sein könne, einen entsprechenden religiösen Glauben zu teilen, führt die Bekundung, dieser sei nur Eingeweihten zugänglich, bei vielen zu Verärgerung, weil sie sich dadurch von vornherein ausgeschlossen fühlen, darüber rational sprechen zu können. Die angesprochene Bekundung führt andererseits zu Frustration, weil sie den Verdacht erweckt, religiöse Menschen könnten ihren Glauben nicht rechtfertigen und seinen somit *intellektuell unredlich*. Zuweilen wird der Verdacht auch explizit ausgesprochen und der entsprechende Vorwurf erhoben.

2. Prämissen

Ich setze hier voraus, daß jene *zwei* philosophischen Positionen, die in der Regel mit dem Frühen Wittgenstein einerseits und dem Späten andererseits in Verbindung gebracht werden, uns in der Frage nach der Rationalität des Glaubens nicht weiterführen. Der eine Standpunkt besteht letztlich in der *Leugnung* jeglicher Form von Rationalität auf dem Gebiet des Religiösen, der andere in der Hervorhebung ihrer *Vielfalt* und Relativität. Beide wurden und werden nicht nur von Atheisten und Agnostikern, sondern auch von Theisten vertreten. Denken wir an die alte Tradition der negativen Theologie oder an die Theologie der Reformatoren. Gott sei zu erhaben, als daß wir Menschen über ihn rational kontrollierbar sprechen könnten.

B. Brogaard, B. Smith (Hg.), *Rationality and Irrationality / Rationalität und Irrationalität*, S. 294–300.

Ein Grund, weshalb ich hier voraussetze, daß die *erste* Position nicht angebracht ist, ergibt sich aus der Schwierigkeit, auf dem Gebiet des Religiösen eine klare Trennung zwischen theoretischer und praktischer Rationalität ziehen zu wollen. Unbestritten ist es nämlich, daß die durch die religiöse Rede ausgedrückten oder vorausgesetzten Überzeugungen Folgen für das menschliche Zusammenleben und die Strukturierung der menschlichen Gesellschaft zeitigen. Als Bestätigung dafür möge der Verweis genügen auf die wiederholten Bemühungen der staatlichen Gewalt, das Religiöse zu kontrollieren. Selbst die heute weit verbreitete Ansicht und die entsprechende Bekundung, Religion sei *Privatsache*, kann als Zeichen des Bemühens gedeutet werden, jene religiösen Einstellungen zu neutralisieren, die sich auf die Öffentlichkeit auswirken könnten. Trotz Hervorhebung der Toleranz werden und wurden politische Instanzen gerade wegen der Überzeugung immer wieder aktiv, religiöse Einstellungen zeitigten entweder schädliche oder vorteilhafte Folgen für das menschliche Zusammenleben.

Ein Grund, weshalb ich hier voraussetze, auch die *zweite* Position sei nicht angebracht, derzufolge es – formelhaft ausgedrückt – so viele Formen von Rationalität wie Sprachspiele gibt, ist, daß man so der Rationalitätsfrage letztlich ausweicht. Ist man überzeugt, daß das in der einen Sprache Behauptete wahr oder zutreffend ist, so muß geklärt werden, wie es zusammenhängt mit dem in der anderen Sprache als wahr Behaupteten. Die Klärung ist besonders dann vonnöten, wenn das so Behauptete den Anschein erweckt, widersprüchlich zu sein. Der Grund entspricht jenem, der u.a. in den Beiträgen der neueren Philosophy of Mind gegen den sogenannten *Sprach-Dualismus* vorgebracht wurde: Zu einfach ist es, das Leib-Seele-Problem so lösen zu wollen, daß man die mentale Alltagssprache in ihrem Eigenrecht neben der positiv wissenschaftlichen Sprache beläßt, ohne zu klären, wie beide miteinander zusammenhängen. Dementsprechend ist es auch zu einfach, lediglich zu bekunden, der religiösen Rede käme eine ihr eigene Form von Rationalität zu, ohne zu klären, in welchem Verhältnis sie zur Rationalität der sonstigen menschlichen Rede stehe.

Die zwei hier genannten Optionen eignen sich nicht, einer zufriedenstellenden Lösung des angeschnittenen Problems näher zu kommen. Wer sie vertritt, weicht diesem Problem aus und löst u.U. ebenfalls Verärgerung bzw. Frustration aus.

In jüngster Zeit haben sich zur genannten Problematik eine Reihe von kalvinistisch geprägten Denkern in den USA geäußert. Ihre „reformed epistemology" hat auch unabhängig von der religiösen Rationalitätsfrage Beachtung gefunden, weil sie von allgemeinem erkenntnistheoretischem Interesse ist.

3. „Reformed Epistemology"

Besonders in der kalvinistischen Tradition wird der Gnadencharakter des christlichen Glaubens hervorgehoben. Der religiöse Glaube wird als Geschenk an die von Gott *Erwählten* angesehen. Fast frevelhaft mutet es also an, meinte man, man könne von sich aus, aufgrund rationaler Überlegung und Argumentation, den Glauben erlangen, begründen oder rechtfertigen. So schreibt z.B. Alston: „Gott ist nicht für Voyeure zu haben. Das Ge-

wahren Gottes und ein Verständnis seiner Natur und seiner Anweisungen an uns ist keine rein kognitive Leistung ..." (Alston 1998, 315) Der Sache nach kommt also die kalvinistische Position jener hier erwähnten der Eingeweihten sehr nahe.

Der für uns relevante Standpunkt dieser calvinistisch geprägten Gruppe von Epistemologen lautet nun, daß zentrale Glaubensannahmen weder durch Argumente gestützt noch auf andere zurückgeführt werden können. Wer sie annimmt oder teilt, sei aber deshalb nicht irrational. Kann also ein gläubiger Christ seine zentralen Glaubenspropositionen nicht rational begründen, kann er sie weder auf andere zurückführen noch durch Sinneserfahrung belegen, dürfe man ihn nicht der intellektuellen Unredlichkeit bezichtigen.

Von alters her ist epistemologisch klar, daß nicht jede Proposition begründet werden kann. Gewisse Annahmen und Einsichten sind *grundlegend*, bilden gleichsam die Basis für andere, die auf sie zurückgeführt oder von ihnen abgeleitet werden. Das galt als selbstverständlich im Kontext der scholastischen Philosophien sowie der positivistisch-empiristischen Konstitutionssysteme. Worin sich die Geister scheiden, ist aber die Frage, welche Annahmen zu Recht zur Basis zu rechnen, welche m.a.W. „*properly basic*" seien.

Besonders Plantinga verteidigt die Ansicht, daß auch religiöse Glaubensannahmen *basal* sind oder sein können und daß es somit durchaus rational sein kann, sie zu teilen. Er plädiert für eine Erweiterung der seit alters als *basal* geltenden Gruppe von Annahmen, nämlich der evidenten (self-evident), der unkorrigierbaren und jener Annahmen, die sich unmittelbar aus der Sinneserfahrung ergeben (im Wiener Kreis wurden diese auch „Beobachtungssätze" genannt). Die Klasse der berechtigterweise basalen Propositionen sei eben größer als gemeinhin angenommen.

Als typisch berechtigterweise basal gelten Überzeugungen wie „Ich sehe einen Baum", aber – so hebt Plantinga hervor – faktisch auch Überzeugungen wie „Ich habe heute morgen gefrühstückt," oder „Diese Person ist zornig" (Plantinga 1998, 322). Zu den „properly basic beliefs" werden also faktisch auch solche gerechnet, die zwar nicht für die Sinne, wohl aber erfahrungsmäßig - in einem umfassenderen Sinn - evident sind. Was sie zu berechtigterweise basalen Überzeugungen macht, ist jeweils ein Umstand oder eine Voraussetzung, der bzw. die als *Grundlage* ihrer Rechtfertigung dient. Ist die entsprechende Bedingung nicht gegeben, darf die jeweilige Überzeugung nicht als basal betrachtet werden. Weiß ich z.B., daß ich eine meinen Blick verzerrende Brille trage, krank bin oder eine Droge eingenommen habe, so sind die geforderten Bedingungen im Fall der genannten Beispiele nicht erfüllt. „Der zentrale Punkt ist hier ..., daß eine Überzeugung nur unter bestimmten Bedingungen berechtigterweise basal ist; diese Bedingungen sind, wie wir sagen könnten, die Grundlage für die Überzeugung selbst. In diesem Sinne sind basale Überzeugungen nicht – oder nicht notwendigerweise – *grundlose* Überzeugungen". (Plantinga 1998, 323)

Warum sollten wir nun von vornherein bestimmte Erfahrungen, die nicht unter die normalen Sinneserfahrungen oder Perzeptionen subsumierbar sind, als Grundlage für religiöse Überzeugungen ausschließen? Plantinga bemüht sich, die Ansicht zu verteidigen, daß die Bandbreite möglicher Erfahrungen, die basale Überzeugungen rechtfertigen, größer ist, als von Empiristen angenommen. Er meint: „Es gibt somit viele Bedingungen und Umstände, die den Glauben an Gott hervorbringen können: Schuld, Dankbarkeit, Gefahr, ein Gefühl für Gottes Gegenwärtigkeit, ein Gefühl, daß er spricht, und die Wahrnehmung

von verschiedenen Teilen des Universums. Eine vollständige Arbeit würde die Phänomenologie all dieser und weiterer Umstände erforschen." (Plantinga 1998, 324)

Daß also religiöse Überzeugungen berechtigterweise basal sein können, daß man sie also rational gerechtfertigterweise annehmen kann, macht den Kern und auch den *Zankapfel* der Reformierten Epistemologie aus. Es hängt davon ab, wie man die Umstände und die Bedingungen der religiösen Erfahrung deutet, will man nicht der Beliebigkeit Tür und Tor öffnen. Eine dieser Bedingungen ist nach Plantingas neueren Schriften z.B. auch, daß die Fähigkeiten oder Vermögen, durch die wir unsere Überzeugungen bilden, normal und gut funktionieren (Plantinga 1993, 99).

Nicht alle Überzeugungen, die auf Erfahrung, auch nicht auf Sinneserfahrung, beruhen, sind – wie wir gesehen haben – tatsächlich berechtigterweise basal. Sie können es zwar *prima facie*, müssen es aber nicht wirklich sein. Ihre Rechtfertigung kann durch sogenannte „*defeaters*", Widerlegungsinstanzen, außer Kraft gesetzt werden. Mir scheint z.B., ich sähe einen Baum, entdecke aber, daß ich vor einer Attrappe stehe; mir scheint, ich hätte es mit einem Menschen zu tun, entdecke aber, daß ich einen Roboter vor mir habe; ich meine, mich erinnern zu können, werde mir aber bewußt, daß ich unter den Folgen einer Droge leide und daß meine kognitiven Vermögen daher nicht richtig funktionieren. Berücksichtigt werden muß ferner, daß Basisüberzeugungen *personen-* und *zeitrelativ* sind. Was für den einen als grundlegend gilt, kann für den anderen als abgeleitet gelten, und für einen selbst kann zu dem einen Zeitpunkt eine Überzeugung basal sein, die es zu einem anderen nicht mehr ist.

Was für basale Beobachtungs- und Erinnerungsüberzeugungen gilt, gilt unter dieser Rücksicht auch für religiöse Überzeugungen: Ihre *prima facie*-Rechtfertigung kann außer Kraft gesetzt werden. Wird z.B. einer Person nachgewiesen, daß ihre vermeintlich religiösen Erfahrungen Halluzinationen sind, auf Verletzung von Organen oder gestörte kognitive Funktionen zurückzuführen sind, so können sie nicht berechtigterweise als Grundlage für religiöse Überzeugungen angesehen werden: Sie sind eben „defeated".

Aber auch darin kann es sich seinerseits nur um eine *prima facie*-Widerlegung handeln. Die erwähnten „defeaters" können nämlich ihrerseits von weiteren „defeaters" außer Kraft gesetzt werden wie im Fall sonstiger Widerlegungsinstanzen von vermeintlich berechtigterweise basalen Überzeugungen. Es könnte sich z.B. herausstellen, daß die vermeintliche Attrappe gar keine ist, sondern doch ein Baum oder daß ich gar nicht krank bin, sondern sehr wohl gesunde Funktionen habe. Und so spricht man von „*defeaters defeaters*".

Widerlegungsinstanzen problematischer religiöser Überzeugungen stammen zumeist von außen, können aber auch von innen kommen. Sie können auch von der gläubigen Person selbst oder ihrer Glaubensgemeinschaft ausgehen. Jede christliche Glaubensgemeinschaft kennt derartige Instanzen, besonders dann, wenn sie eine gewisse Größe hat und strukturell organisiert ist.

Die *dritte* von mir hier erwähnte *Position* im Umgang mit der angeschnittenen Problematik des Glaubens der „Eingeweihten" nimmt also an, daß Glaubensüberzeugungen, die auf Erfahrungen und auf eine Praxis gründen, die nicht von jedermann gemacht bzw. geteilt werden, insofern rational sein können, als sie berechtigterweise basal sind. Sie kommt also dem Standpunkt jener entgegen, die ihren Glauben weder auf andere Propositionen zurückführen noch durch Argumente zu stützen versuchen, sondern lediglich auf

ihre Erfahrungen oder die Praxis im Rahmen der Glaubensgemeinschaft als Grundlage ihrer Überzeugungen verweisen. Durch diese Position wird aber nicht ausgeschlossen, daß Menschen, die sich weigern, auf sogenannte „defeaters" ihrer basalen Annahmen zu reagieren, insofern nicht rational sind, als sie trotz Widerlegungen ihrer Glaubensannahmen gleichgültig bleiben. Finden sie allerdings Widerlegungen dieser Widerlegungen, sogenannte „defeaters defeaters", können sie durchaus als rational gelten.

Im Rahmen der Reformed Epistemology. wird nicht vorausgesetzt, daß Rationalitätsfragen nur dann zufriedenstellend gelöst werden können, wenn es gelänge, objektive, für alle bindende Rationalitätskriterien zu finden oder zu entwickeln. Die Praxis der „defeaters" und der „defeaters defeaters" ist auch ohne derartige Kriterien möglich. Es sei zwar wünschenswert, derartige Kriterien zu entwickeln, diese könnten aber nicht a priori dekretiert werden, sondern ergäben sich höchstens induktiv. Wie das zu erfolgen habe, ist aber äußerst umstritten. (Siehe die Diskussion mit Quinn und Gutting.)

Zuletzt möchte ich eine weitere Position erwähnen, die ich allerdings nur von ihrem Anspruch her umschreibe, nämlich die der Römisch-Katholischen Kirche.

4. Römisch-Katholische Position

Zu unserer hier angesprochenen Problematik steht die Römisch-Katholische Kirche in einem *Spannungsverhältnis*. Auf der einen Seite teilt sie die Einstellung, daß zentrale Glaubensinhalte und entsprechende rituelle oder sakramentale Vollzüge nur den Eingeweihten bzw. Getauften zugänglich sind und letztlich weder durch Argumentation noch durch sonstige rationale Leistungen erworben werden können, sondern *Geschenk-* oder *Gnadencharakter* haben. Auf der anderen Seite teilt die Kirche die Überzeugung, daß die Glaubensinhalte nur auf Grund der Überzeugung, daß sie *wahr* sind, angenommen werden können. Sind sie aber wahr, so können sie nicht im Widerspruch zu den anderen als wahr geglaubten oder angenommenen Überzeugungen stehen. Diese Position hat ihren Niederschlag in offiziellen kirchlichen Verlautbarungen und konziliären Dekreten gefunden.

In der christlichen Tradition finden wir zwar von Anfang an rationalitätsfeindliche Einstellungen, diese konnten sich aber auf der Ebene der organisierten offiziellen Kirche nie ganz durchsetzen. Wir finden sie bereits – wenn auch z.T. nur implizit – in den Paulinischen Schriften, so z.B. in 1 Kor. 1: Was die Welt für töricht hält, hat Gott erwählt, um die Weisen und Klugen dieser Welt zu beschämen! Besonders in den Schriften der Kirchenväter können wir rationalitätsfeindliche Ansichten eruieren. Geistesgeschichtlich relevant wurde z.B. Tertullians Diktum „credo quia absurdum".

Sowohl aus praktischen wie auch theoretischen Gründen wurde aber der andere Pol des Spannungsverhältnisses – zumindest vom Anspruch her – nie vernachlässigt. Selbst in den düsteren Zeiten des Antimodernismusstreits um die Jahrhundertwende hat die offizielle Kirche den entsprechenden Standpunkt nicht verlassen, sondern im Gegenteil explizit hervorgehoben. Christliche Glaubensinhalte und Vernunfwahrheiten könnten einander nicht ausschließen. Kurz auf eine Formel gebracht: Christlicher Glaube und Vernunft können einander nicht widersprechen. (Vaticanum I) Und trotz der vielen Mängel, die von philosophischer Seite gegen die Enzyklika „Fides et Ratio" von Johannes

Paul II. (15.10.1998) vorgebracht werden, ist klar, daß gerade sie ein starkes und in der Römisch-Katholischen Tradition tief verankertes Bekenntnis zur Rationalität des christlichen Glaubens enthält.

Die Römisch-Katholische Kirche kennt zudem ein strenges inneres Verfahren, das man in der Terminologie der Reformed Epistemology als „defeater-Verfahren" bezeichnen könnte, um rein private Erleuchtungen und Offenbarungen in ihrer Relevanz und Nachvollziehbarkeit für die größere Gemeinschaft sowie auf Rationalität hin zu überprüfen und notfalls zurückzuweisen.

Die angesprochene Problematik wird kirchenintern auch unter dem Stichwort „*Mystagogie*" behandelt, welches eine begleitete Hinführung zu den Glaubensmysterien bedeutet, die zunächst nicht jedermann zugänglich sind. Wer nicht hingeführt und eingeweiht wird, kann sie nicht verstehen. Daraus folgt aber nicht, daß die Annahme der Inhalte – zumindest vom Anspruch her – irrational wäre. Dem Mystagogen kommt die Aufgabe zu, den Neophyten oder Neugetauften einzuführen, sodaß er verstehen und erfassen kann, was ihm zuteil wird.

5. Ausblick

Wir haben eingangs gesehen, daß Konflikte und Auseinandersetzungen über die Rationalität von religiösen Überzeugungen – speziell jener, die lediglich „Eingeweihten" zugänglich sind – häufig zu Verärgerung einerseits und Frustration andererseits führen. Sie lassen die Kontrahenten auch deshalb nicht neutral, weil sie *Folgen* zeitigen für die Gestaltung und Bewältigung des gesellschaftlichen Lebens.

Diese Verzahnung zwischen theoretischer und praktischer Dimension der Rationalitätsfrage von religiösen Überzeugungen vereitelt Lösungen im Sinne rein theoretischer einerseits und rein praktischer Rationalität andererseits. Religiöse Überzeugungen sind in ihrer Komplexität eher so einzuordnen wie *weltanschauliche* Überzeugungen.

Mit „*Weltanschauung*" ist nicht primär ein theoretischer Entwurf oder eine Theorie über die Welt gemeint, sondern ein komplexes Gebilde von Überzeugungen und Einstellungen, die Lebensentscheidungen bestimmen und von solchen abhängig sind. Darunter fallen somit auch die persönlichen und gelebten Haltungen, aus denen heraus man das im Alltag Begegnende spontan auffaßt, *einordnet* und *bewertet*. Besonders insofern sie einen persönlichen und praktischen Aspekt aufweisen, lassen Weltanschauungen Menschen nicht gleichgültig.

Obwohl nun weltanschauliche Sätze nicht wie Sätze von Theorien oder wie rein empirische Aussagen einzuordnen sind, insofern sie eher Einstellungen zur Wirklichkeit zum Ausdruck bringen, aus denen heraus man lebt, zeigen sie dennoch eine *rationale Struktur*, wenn auch eigener Art. In ihrer Rationalität unterscheiden sie sich zwar von Theorien, kommen aber auch unter verschiedenen Rücksichten mit ihnen überein, so müssen sie z.B. den Kriterien der *Widerspruchsfreiheit* und *Einheitlichkeit* genügen. Gerade insofern sie Weltanschauungen sind, müssen sie aber auch *umfassend* sein und dürfen nicht von vornherein bestimmte Lebensbereiche und Erfahrungen ausschließen. Diese Kriterien „machen verständlich sowohl das rationale Element der Weiterentfaltung oder Änderung weltanschaulicher Überzeugungen wie auch die Möglichkeit, daß trotz des persönli-

chen Charakters weltanschaulicher Überzeugungen darüber interpersonal argumentiert werden kann, wenigstens um die Erfüllung der Kriterien zu überprüfen." (Muck 1999, 133)

Aus der Eigenart von Weltanschauungen folgt auch, daß die Rationalität religiöser Überzeugungen weder an der Möglichkeit gemessen werden darf, einen Konsens herzustellen, noch an der Vergleichbarkeit mit der Rationalität von Wissenschaft: „Darum ist hier auch nicht die Sicherheit zu erwarten, die man in naturwissenschaftlicher Erkenntnis gewohnt ist. Wohl aber gibt es vergleichbare *Prüfungsmöglichkeiten*. So kann gefragt werden, ob tatsächlich die gesamte Lebenserfahrung berücksichtigt wird und ob auch neue Lebenserfahrung einbezogen werden kann." (Muck 1999, 134)

Es sollte also vor dem Hintergrund der Römisch-Katholischen Kirche sowie des im Kontext der Reformed Epistemology vertretenen Standpunktes der *properly basic beliefs* durchaus möglich sein, über religiöse Glaubensannahmen rational zu argumentieren. Wie in weltanschaulichen Auseinandersetzungen ist es aber notgedrungen schwer, wenn nicht unmöglich, einen für die beteiligten Parteien überzeugenden Konsens und eine entsprechende Sicherheit zu finden. Die Rationalitätsstandards des religiösen Glaubens weichen nämlich von den theoretischen der positiven Wissenschaften wegen ihrer unterschiedlichen Funktionen ab. Die Bekundung von religiös Eingeweihten, ihre Glaubensinhalte seien nur den Eingeweihten zugänglich, mag zwar unter bestimmten Rücksichten stimmen, darf aber nicht als Immunisierungsmittel mißbraucht werden, um sich dauerhaft der rationalen religiös-weltanschaulichen Auseinandersetzung zu entziehen oder auf mögliche „defeaters" ihrer religiösen Überzeugungen nicht zu reagieren.

Literatur

Alston, W.P. 1998 „Religiöse Erfahrung und religiöse Überzeugungen", in: Ch. Jäger (Hg.), *Analytische Religionsphilosophie*. Paderborn: Schöningh, 303–316.

Muck, O. 1999 *Rationalität und Weltanschauung*. Philosophische Untersuchungen, Innsbruck/Wien: Tyrolia.

Plantinga, A. 1993 *Warrant and proper function*, Oxford: University Press.

Plantinga, A. 1998 „Ist der Glaube an Gott berechtigterweise basal?", in: Ch. Jäger (Hg.), *Analytische Religionsphilosophie*. Paderborn: Schöningh, 317–330.

Runggaldier, E. 1996 *Philosophie der Esoterik*, Stuttgart: Kohlhammer.

Swinburne, R. 1981 *Faith and Reason*, Oxford: Clarendon Press.

Kinds of Rationality and their Role in Evolution

GERHARD SCHURZ

1. The Concept of Rationality: Some Meaning Components

1.1 Objects of Rationality

Let us first clarify to which kind of objects the predicate of rationality applies. Prima facie, it is *actions* which can be judged as being more or less rational. Or better, action-generating strategies, modes of behaviour, because no action can be evaluated with respect to its rationality in isolation from other actions.

Conscious actions are guided by descriptive beliefs and purposes, and consciously reflected purposes are themselves expressed in terms of beliefs, namely normative or evaluative beliefs. So at a second level, which is the philosophically more important one, it is beliefs which can be judged as being more or less rational. Or better, *belief systems* including cognitive methods – because again, beliefs cannot be evaluated in isolation from other beliefs, and in isolation from the cognitive methods from which they result. Summarized, the objects of rationality are prima facie modes of behavior, and at a deeper level, belief systems. So much about the objects of the rationality predicate.

1.2 Three meaning postulates

What does it *mean*, for a mode of behaviour or belief system, to be *rational?* In the following, I want to ask this question at a philosophically *fundamental* level, without presupposing any fixed conception of rationality such as scientific rationality. Thus, I have to start my analysis from a sort of philosophical vacuum. The best we can do at this stage is to look for meaning 'postulates' (cf. Schurz 1997, ch. 11), which give us some core components of rationality, on which *normal* speakers of ordinary and philosophical discourse agree.

The first meaning component of rationality is that it is not a purely descriptive concept: it has a *normative* element. We cannot say: a man's principles of action have been completely rational *and* the consequences for which his actions were responsible were overwhelmingly bad. Being rational implies to have overwhelmingly good consequences, at least *ceteris paribus* and under *normal* circumstances.[1] I admit that this is a strong restriction, but I think it is a fair one. Whatever your definition of rationality is, if I can convincingly demonstrate that acting according to your definition has overwhelmingly bad consequences; then (almost) everyone would agree that your definition has *failed* to capture rationality.

However, rationality does not simply coincide with the good. The second meaning component of rationality is its epistemic component. To be rational means to have good

1. This formulation also shows that laws of rational behavior are not strict but *normic* (cf. Schurz 2001a).

B. Brogaard, B. Smith (Hg.), *Rationality and Irrationality / Rationalität und Irrationalität*, S. 301–310.

consequences by way of *reason*, or *intellect*. The instinctive behavior of the epistemologically savage or primitive man may be good, but it is not rational. In other words, if something is rational, then it is justified by reason that it has some good consequences. In this respect, rational action or belief is related to the good like knowledge is related to truth.

But there is a third aspect: rationality is relative to some given *purposes*. If we call an action rational, we have a purpose in mind with respect to which the action is rational. Now, is rationality really completely purpose-relative? At this stage of the discussion we already enter the major problem which I want to discuss in this paper. Of course, we may define a purely *instrumental* concept of rationality, by calling anything rational with respect to a given set of purposes which is an optimal means to achieve these purposes. But the so-defined concept would be a completely *opportunistic* one, not worth of being called a concept of rationality; in particular because it is not clear whether these purpose-relative instrumental 'rationalities' have anything essential in common. Still, our cultural tradition has something which is in perfect fit with this instrumental and purpose-relative aspect of rationality – the so-called theoretical rationality. With the only difference that theoretical rationality it is not characterized pragmatically, in terms of purpose-relative success, but in terms of truth.

2. Enlightenment Rationality and its Alternatives

2.1 Theoretical and practical rationality

We thus arrive at the central rationality concept of our own, western cultural tradition: the tradition of *enlightenment*. To avoid misunderstandings, with enlightenment I don't mean particular philosophical traditions like rationalism versus empiricism, but I mean the whole tradition of rational reflection, which starts in Antiquity and has been continued, after a period of stagnation in the middle ages, in the late scholastic period, in the classical enlightenment philosophy of Descartes, Hume, and Kant, and which has culminated in the modern system of sciences. The main alternatives to the enlightenment tradition are, as we shall see, all sorts of religious and esoteric worldviews. According to the enlightenment tradition, rationality rests on the two columns of theoretical and practical rationality.

> *Theoretical rationality:* A belief system is theoretically rational iff [or, the more] its beliefs tend to be true and its methods are optimal means to achieve true beliefs, in the correspondence-theoretic and scientific sense of truth.

To avoid misunderstanding: I shall here explicitly focus on the correspondence-theoretic notion of truth, more specifically on its scientific understanding related to the criteria of empirical adequacy, etc. Of course, there is an ambiguity in the common sense notion of truth: it may be used as equivocally with rationality. For example, one who conceives religion as a *deeper* form of reason will also be inclined to call religion as a deeper and non-scientific form of truth. I do not want to use the notion of truth in this more general and very unclear sense, because I do not think that this leads us anywhere. Hence, I shall

focus on the correspondence-theoretic and scientific concept of truth. I assume that a reasonable characterisation of theoretical and scientific rationality is possible (cf. Schurz 1998), but I will not discuss any questions of this sort here. My point will lie in a completely different direction. However, one characteristic of theoretical rationality will be of crucial importance for my problem: it is its *intrinsic self-control* in the form of openness to criticism (a fact which is particularly emphasized in the work of Popper).

The major problem of theoretical rationality is that it is a purely descriptive concept: it has lost its normative component. It is a commonplace that knowledge may be utilized for good as well as for very bad purposes. Even a mass murderer, a kind of Hitler, can be a theoretically rational being. As a matter of fact, the *main* evolutionary consequence of the development of theoretical and scientific rationality was not the increase of humanity and wisdom, but the increase of technology, world population, and consumption. These are the facts, and we philosophers have to face them.

It is not of much help here to pray to the public that scientific truth is the utmost intrinsic value, which is not in need of further justification. This may be so for us academic people, but certainly not for the high majority of people. Happiness and well-being, mental, emotional and social harmony are the factual intrinsic values of man as a natural being, but not truth – at least if we understand this concept in a precise scientific sense.

Theoretical rationality in the scientific sense is a purely descriptive concept; it does not have an intrinsic connection with the good. Therefore, it can impossibly be *all* there is to rationality. On this reason, the tradition of enlightenment rationality has always postulated a second column on which rationality is based, and this is practical rationality. Practical rationality is not purpose-relative; it assumes certain values to be fundamental for the human condition:

> *Practical rationality:* A mode of behavior or a belief system is practically rational, iff [or, the more] it contributes to the realization of basic values of humanity such as: preservation of life, happiness and well-being, social cooperation and non-violence.

It is a matter of ethical debate which values are more fundamental than others, and whether fundamental values can be justified in as strong a way as scientific truths. But again, this question is not important for my problem. It is sufficient for me to assume that most people and cultures would agree on the above values, or at least that they have been driven by cultural evolution to agree on these values. My point lies in a different direction, namely the following.

2.2 The enlightenment thesis

It is a core thesis of enlightenment rationality that the search for theoretical truth is the canonical way to achieve the goals of practical rationality. It was the central intention of the enlightenment tradition, as opposed to earlier worldviews which were based on religious authority, to practically improve the world and the human condition *by gaining theoretical understanding and knowledge about it*. I summarize this in the following thesis of enlightenment rationality:

Thesis of enlightenment rationality: The optimal means for achieving a given set of purposes – in particular, for achieving the purposes of practical rationality – is to act according to a theoretically rational belief system.

This thesis explains *how* theoretical and practical rationality are intended to cooperate in the project of enlightenment rationality. Practical rationality tells you the right purposes, and theoretical rationality tells you the optimal means to reach them. The ability of theoretical rationality to tell you the optimal means is, or course, not restricted to particular purposes, but may be utilized for any set of purposes. In other words, the thesis of enlightenment rationality does not say that theoretical rationality leads *by its own* to good consequences – we already know that even a Hitler may be theoretically rational – but rather, that theoretical rationality if guided by the 'right' purposes will lead to good consequences.

2.3 Religious and esoteric world-views

Here comes the main question of my paper: is the enlightenment thesis true? Is there a systematic connection between knowledge of the truth and maximizing success in practical purposes?[2]

What would be an alternative to enlightenment rationality? For example, take recent psychological trends to establish new non-cognitive forms of intelligence, for example 'emotional intelligence' (Goleman 1995), as a way of being successful in social communication and power management. Emotional intelligence is not necessarily an alternative to enlightenment rationality. It can equally be regarded as a rational psycho- or socio-technical instrument to achieve success in business, politics or whatever. More generally, every system of behavioural abilities which is successful with respect to practical purposes is embeddable into enlightenment rationality iff the following two conditions are fulfilled:

1. Its success is theoretically explainable.
2. Learning and performing these abilities is compatible with having a theoretically rational belief system.

For example, the defender of a scientific worldview, after having red Goleman's book, can easily say to him- or herself that, on rational grounds, some training in emotional intelligence would be good for his or her practical success.

Since I believe in theoretical rationality, I do not believe in the existence of practical success which, on reasons of principle, has no theoretical explanation – which would be a violation of condition 1. So, the only serious alternative to enlightenment rationality are those belief systems or practices which violate condition 2. These are exactly the religious and esoteric world-views in a generalized sense. The success of these world views requires that one has false beliefs, or to have beliefs which are not open to criticism, i.e., which are not allowed to be scientifically tested:

Central components of religious (esoteric) rationality: 1. Belief in supernatural beings (God) or powers (spirits, energies, fields, charisma, etc.) which are incompatible with scientific knowledge. 2. Belief as direct means to practical success under exclusion of the method of critical test: you must believe in the *first* place, in order to get convinced of your belief.

2. This question is crucial for pragmatic philosophies; cf. Schurz 1998.

Again, a remark is necessary to avoid misunderstanding. Of course, one may define a philosophically abstract concept of God and rationally discuss its logical properties. But I refer to the religious and esoteric worldviews as they have really played their role in the history of mankind, with all of their characteristics. In this history, views which were connected with religion have been refuted by science again and again – be it the geocentric world view, be it wrong views about evolution, be it views in prophecy and miracles, or be it morally unacceptable views, e.g. that of the superiority of man over woman. Thus, these religious and esoteric worldviews are everything else but theoretically rational.

3. Evolution of Rationality

3.1 Evolutionary epistemology and the facts of evolution

Evolutionary epistemology is based on the assumption that those belief systems which have been successful in evolution are, in the long run, also those which are closest to the truth, or (if they are methods) which are most truth-conducive. Hence, evolutionary epistemology is just a modernized form of the enlightenment thesis. I come now to the major challenge of my paper. I will argue that if the enlightenment thesis, in the form of evolutionary epistemology, were generally true, then religious worldviews should have been eliminated in the process of cultural evolution. Since this was not so, there must be somewhere a mistake. Let me elaborate my point.

3.2 Generalized evolution theory

I understand 'evolution' in the sense of generalized evolution theory. It was suggested as a possibility by Dawkins in his conception of 'memes' and has been theoretically developed by Boyd and Richerson (1985). In distinction to sociobiology, general evolution theory assumes cultural evolution as an independent level of evolution which is not determined by genetic evolution. Yet, cultural evolution theory shares with genetic evolution three modules (cf. Schurz 2001b):

(E1) A mechanism of *variation* which acts in larger populations of mutually competing systems. In our case, these systems are competing worldviews and systems of rationality, which are created by human inventiveness and creativity.

(E2) A mechanism of *reproduction*. In the case of cultural evolution, this is the tradition of knowledge and ideas from one generation to the next, in the form of socialization and education.

(E3) Finally, an environment which *selects* the fittest among the variations, i.e., those with the highest cultural reproduction rate. In cultural evolution, selection is performed by the society, by way of complex procedures which depend on the political system. In democratic systems, where freedom of opinion is an institutionalised right, selection takes place within a 'market of ideas'.

The paradigm example which illustrates the enormous power of cultural evolution is the evolution of science and technology. The knowledge accumulated in this area exceeds what can be learned in an individual's lifetime by millions. Evolution in this area has ruthlessly eliminated all technological inventions of earlier times because of their minor success in the technological market. Cultural evolution has also ruthlessly eliminated social habits or social structures of earlier times because of their minor reproduction rate. Thus, cultural evolution was indeed enormously effective and powerful. Why is it that in spite of these facts religious worldviews are still so common, even in the U.S.A. as the foremost part of high tech societies?

Sociobiologists try to explain this fact by the assumption of a certain genetic disposition for religion (Wilson 1998, Wenegrat 1990). Although that might be true, it does not really solve our problem. For as a matter of fact, one is not genetically forced into belief in God; one *can* live a successful life without a religious worldview. So man *has* a cultural choice here. And since religious and esoteric world views are predominantly false, and moreover irrational because of their exclusion of the method of critical test, the consequence is unavoidable that they should have been eliminated by cultural evolution, *if* the thesis of enlightenment rationality, namely that theoretical rationality is the optimal means to achieve practical rationality, were true. Where is the mistake?

4. The Blind Spot of Enlightenment Rationality: the Generalized Placebo-Effect

4.1 Truth-effects versus placebo-effects of beliefs

A crucial assumption of the enlightenment thesis is that the practical effects which our beliefs have upon us are predominantly determined by the question of the truth or truthlikeness of our beliefs. For example, if I believe that it will be rainy tomorrow, then the practical effect of this belief on me, namely that I take an umbrella with me tomorrow, is determined by the question of its truth: if my belief is true, then the effect will be that I stay dry, which is good, and if my belief is false, then the effect will be that I have to carry a superfluous umbrella with me, which is bad. I call these kinds of effects the *truth-effects* of beliefs. In other words, the crucial assumption of enlightenment rationality and of evolutionary epistemology is that all important effects of belief systems are truth-effects.

However, if we regard man as a natural living being then it is obvious that, in general, beliefs will have many effects which do *not* depend on the truth of the beliefs. For example, if I believe that my girl friend will visit me in one hour, then this makes me happy during the next hour, completely independent of the question whether my belief is true. More generally, beliefs may have per se positive or negative consequences on us, independently of their truth. I call these effects *generalized placebo-effects* (there is a lot of research in recent psychology on this topic; cf., e.g., Harrington 1997). Figure 1 illustrates the situation.

Thus, I argue that the mistake of the enlightenment thesis lies in the neglect of the placebo-effects of our belief systems.

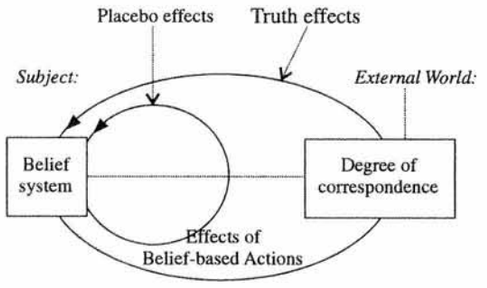


Figure 1: Truth- versus placebo-effects of beliefs

4.2 Illustrations of placebo-effects.

Let me illustrate some kinds of placebo effects: 1. The placebo effect was originally studied in *pharmacy* as the effect of placebo pills, i.e. pills without any pharmacological effects. Placebo effects, for example that of placebo sleeping pills, may be very strong; 50% as strong as the real pharmakas or even stronger. It is also well documented that *psychotherapies* rely to a high degree on the placebo effect.

2. Our modern system of *advertisement* and *propaganda* in economy is to a high degree based on the placebo effect. Self-fulfilling or self-destroying prophecies are special cases of the generalized placebo effect which have been studied in sociology, e.g. in voting processes, or in economical processes of the stock market. In all these cases, the mere beliefs of people have enormous effects, either economically or politically, independent of the question whether these beliefs are true, or – in the case of self-fulfilling or self-destructing prophecies – whether they were *initially* true.

3. To get closer to our topic of esoterics and religion: cancer patients who strongly believe in their recovery have significantly higher expectations to recover than those who do not. Clearly, such strong beliefs are much better motivated by esoteric views about the spiritual powers than by a scientific belief system. In the same way, various kinds of *esoteric techniques* are handled on the contemporary 'market of ideas' which are designed to improve one's state of health, or one's self-confidence, or one's social 'charisma', etc. But all these esoteric techniques work only because they rely on beliefs in spiritual powers which are scientifically false.

4. In fact, even the psychological rule of *positive thinking*, which has become so fashionable in present time folk psychology, is based on the placebo effect. Positive thinking gives you intrinsic strength and self-confidence. But it is in conflict with a strict orientation of your belief system towards truth, for it tells you that you should always believe that there is a way to reach your goals, even if there is no such way. More generally, one can influence one's own emotional condition much more successfully by exploiting the placebo effects of beliefs than by truth-oriented theoretical reasoning.

5. Examples of the previous sort could be multiplied; but we now turn to our central topic. The two major *placebo effects of religious worldviews* are these:

5.1 Increase of psychological happiness, self-confidence and well-being. To believe in good spiritual powers who not only guide you in the mundane life, but who also will reward you for your pain in an eternal life after your mundane life – this gives you a high degree of self-confidence and positive thinking which enlightenment rationality cannot give you.

5.2 Motivation for social cooperation, which is indispensable for the functioning of complex societies. If you believe that you will be rewarded for your good and socially cooperative behavior by God or by higher spiritual powers, then this is a much better support of social cooperation than scientific truth. For the scientific truth tells you that, although social cooperation is generally important, it is always possible to undermine social cooperation by being parasitic, and if you do so in the right way without being caught by police, then you will profit more than those who strictly obey the rules of social cooperation.

So, increase of happiness, self-confidence and motivation for obedience under social rules are, I think, the two main reasons why religious world-views continue to be successful in cultural evolution, in spite of their conflict with scientific truth.

5. Discussion of consequences

Let me summarize. Placebo-based rationality systems, with religion as their paradigm representative, are the major alternatives to Western enlightenment rationality. In spite of their dangers, to which I will come now, they have practically positive effects, which enlightenment rationality cannot exploit. This explains the evolutionary success of placebo-based worldviews. Let me also summarizes three clarifications to avoid misunderstandings:

1. Placebo effects are not illusionary effects but real.
2. Placebo effects do not disconfirm the truth of the scientific worldview; rather, they can be scientifically explained.
3. Placebo-based rationalities are incompatible with enlightenment rationality *not* because the enlightenment rationalist cannot theoretically explain placebo-effects, but rather, because the enlightenment rationalist cannot *exploit* these placebo effects as long as he insists in orienting his beliefs only on the criterion of truth. For example, we cannot think that God does not exist and at the same time believe in God on the pragmatical reason that this makes us happy; a mind who could do this would have to be schizophrenic.

Let me conclude with some remarks on the normative question: which rationality system is better, which should one choose? This gives me the possibility to avoid further misunderstandings. Of course, I am a defender of theoretical rationality and enlightenment rationality. As I have pointed out, the major practical advantage of enlightenment rationality is its *intrinsic self-control* and *self-corrigibility* because of its openness of the method of critical test. Popper has expressed this by saying that in the scientific method we kill theories instead of people. This openness to criticism is a property which placebo-based rationalities cannot have, because the placebo effect cannot arise if we doubt it and want

to test it before we believe in it. Therefore, whenever social or individual decisions are at stake which involve high risks or dangers, it would be mad to base these decisions on religious worldviews. In fact, today Islamic countries, which in spite of the modern level of technology and military power still base their decisions on religion, are one of the greatest risks of contemporary mankind. But religions are not the only worldviews who carry this danger of uncontrolled irrationalism in them. For example, also fascistic worldviews who believe in the higher destination of certain races or nations are similarly dangerous; and likewise dangerous is blind trust in the almightiness of technology.

So I want to defend enlightenment rationality. But also, I want to cure it from its blind spots. The purpose of my analysis was mainly conceptual and theoretical. I have pointed out, first, that the enlightenment thesis of rationality, and in particular modern evolutionary epistemology, overlooks the placebo-effects of belief systems. And second, that these placebo-effects are the theoretical explanation of the continuous evolutionary success of religious and esoteric belief systems, from the stone age until our modern electronic age. Now, I do not claim that I have especially original normative recommendations as a result of my theoretical analysis. In the contrary, I am somehow puzzled by these result. But if one presses me, I would be inclined to support the following practical consequences:

1. That religious or esoteric worldviews have advantages which enlightenment rationality cannot have does not at all mean that there is a need of an 'alternative' theoretical rationality, as proclaimed by post-modernists or other present day philosophers. All we need for an explanation is the generalized placebo-effect.

2. To refrain from the advantages of placebo-effects is, I think, the unavoidable *price* which we have to pay if we want to enjoy the advantages of theoretical and of enlightenment rationality. I think that these advantages are clearly overwhelming. In particular, enlightenment rationality is the only kind of rationality which should be taught in educational institutions and be supported by political institutions.

3. At the same time, we cannot say that placebo-based worldviews or practices are completely irrational. To make individual lifes endurable, a certain dose of placebo effects seems always to be necessary, and particularly in our modern, emotionally cold and technologically complex societies.

4. In a pluralistic and democratic society, it is impossible to eliminate the cultural influence of religion or esoterics. In particular, it is predicted by cultural evolution theory that attempts to diminish the influence of religion or esoterics by way of theoretical persuasion can have only little effect. Also this is a *price*, namely the price which we have to pay for a democratic and pluralistic society.

5. So we should tolerate placebo-based worldviews or practices in those areas where they have positive or at least not negative effects. But on the same reason, a continuous *ideological watchfulness* is necessary, to defeat all placebo-based worldviews as soon as they show tendencies of totalitarianism or violate principles of humanity.

References

Boyd, R. and Richerson, P. J. 1985 *Culture and the Evolutionary Process*, Chicago: Univ. of Chicago Press.

Goleman, D. 1995 *Emotional Intelligence*, New York: Bantam Books.

Harrington, A. et al. (ed.) 1997 *The Placebo Effect: an Interdisciplinary Exploration*, Harvard: Harvard University Press.

Schurz, G. 1997 *The Is-Ought Problem*, Dordrecht: Kluwer.

Schurz, G. 1998 "Kinds of Pragmatisms and Pragmatic Components of Knowledge", in: P. Weingartner et al. (eds.), *The Role of Pragmatics in Contemporary Philosophy*, Vienna: Hölder-Pichler-Tempsky, 9–22.

Schurz, G. 2001a "Normische Gesetzeshypothesen und die wissenschaftsphilosophische Bedeutung des nichtmonotonen Schliessens", to appear in *Zeitschrift für Allgemeine Wissenschaftstheorie*.

Schurz, G. 2001b "Natürliche und kulturelle Evolution: Skizze einer verallgemeinerten Evolutionstheorie", to appear in: L. Solwiczek, W. Wickler (eds.), *Wie wir die Welt erkennen* (Grenzfragen Bd. 27), Freiburg: Karl Alber Verlag.

Wenegrat, B. 1990 *The Divine Archetype: The Sociobiology and Psychology of Religion*, Lexington Books.

Wilson, E. O. 1998 *Consilience. The Unity of Knowledge*, New York: Alfred A. Knopf.

The Classical Model of Rationality and Its Weaknesses[1]

JOHN R. SEARLE

I. The Classical Model of Rationality

In our intellectual culture, we have a quite specific tradition of discussing rationality and practical reason, rationality in action. This tradition goes back to Aristotle's claim that deliberation is always about means, never about ends, it continues in Hume's famous claim that, "Reason is and ought to be the slave of the passions", and in Kant's claim that, "He who wills the end wills the means". The tradition receives its most sophisticated formulation in contemporary mathematical decision-theory. The tradition is by no means unified, and I would not wish to suggest that Aristotle, Hume, and Kant share the same conception of rationality. On the contrary, there are striking differences between them. But there is a common thread, and I believe that of the classical philosophers, Hume gives the clearest statement of what I will be referring to as "the Classical Model". I have for a long time had doubts about this tradition and I am going to spend most of this essay exposing some of its main features and making a preliminary statement of some of my doubts.

When I first learned about mathematical decision-theory as an undergraduate in Oxford, it seemed to me there was an obvious problem with it: it seems to be a strict consequence of the axioms, that if I value my life and I value twenty five cents (a quarter is not very much money but it is enough to pick up off the sidewalk, for example) there must be some odds at which I would bet my life against a quarter. I thought about it, and I concluded there are no odds at which I would bet my life against a quarter, and if there were, I would not bet my child's life against a quarter. So, over the years, I argued about this with several famous decision-theorists, starting with Jimmy Savage in Ann Arbor and including Isaac Levi in New York, and mostly they came to the conclusion, usually after about half an hour of discussion: "You're just plain irrational". Well, I am not so sure. I think maybe they have a problem with their theory of rationality. Some years later the limitations of this conception of rationality were really brought home to me (and this has some practical importance), during the Vietnam War when I went to visit a friend of mine, who was a high official of the Defense Department, in the Pentagon. I tried to argue him out of the war policy the United States was following, particularly the policy of bombing North Vietnam. He had a Ph.D. in mathematical economics. He went to the blackboard and drew the curves of traditional micro economic analysis; and then said, "Where these two curves intersect, the marginal utility of resisting is equal to the marginal disutility of being bombed. At that point, they have to give up. All we are assuming is that they are rational. All we are assuming is that the enemy is rational!"

1. This essay is an expanded version of a talk which formed the opening session of the Kirchberg Wittgenstein Conference of the year 2000. Much of the material in this essay is included in the first chapter of my forthcoming book, *Rationality in Action*, MIT Press.

B. Brogaard, B. Smith (Hg.), *Rationality and Irrationality / Rationalität und Irrationalität*, S. 311–324.

I knew then that we were in serious trouble, not only in our theory of rationality but in its application in practice. It seems crazy to assume that the decision facing Ho Chih Minh and his colleagues was like a decision to buy a tube of toothpaste, strictly one of maximizing expected utility, but it is not easy to say exactly what is wrong with that assumption. As a preliminary intuitive formulation we can say this much: In human rationality, as opposed to animal rationality, there is a distinction between reasons for action which are entirely matters of satisfying some desire or other and reasons which are desire independent. The basic distinction between different sorts of reasons for action is between those reasons which are matters of what you want to do or what you have to do in order to get what you want, on the one hand, and those reasons which are matters of what you have to do regardless of what you want, on the other hand. Things you have to do regardless of whether you want to do them are typically matters of commitments involving such things as obligations and duties.

II. Six Assumptions Behind The Classical Model

I will begin by stating and discussing six assumptions that are largely constitutive of what I have been calling "The Classical Model of Rationality". I do not wish to suggest that the model is unified in the sense that if one accepts one proposition one is committed to all the others. On the contrary, some authors accept some parts and reject other parts. But I do wish to claim that the model forms a coherent whole, and it is one that I find both implicitly and explicitly influential in contemporary writings. Furthermore, the model articulates a conception of rationality that I was brought up on as a student of economics and moral philosophy in Oxford, and it did not seem to me satisfactory then. It does not seem to me satifactory now.

1. *Actions, where rational, are caused by beliefs and desires.*

Beliefs and desires function both as causes and as reasons for our actions, and rationality is largely a matter of coordinating beliefs and desires so that they cause actions "in the right way".

It is important to emphasize that the sense of "cause" here is the common or Aristotelian "efficient cause" sense of the word where a cause of an event is what makes it happen. Such causes, in a particular context, are sufficient conditions for an event to occur. To say that specific beliefs and desires caused a particular action is like saying that the earthquake caused the building to collapse.

2. *Rationality is a matter of obeying rules, the special rules that make the distinction between rational and irrational thought and behavior.*

Our task as theoreticians is to try to make explicit the inexplicit rules of rationality that fortunately most rational people are able to follow unconsciously. Just as they can speak English without knowing the rules of grammar, or they can speak in prose without knowing that they are speaking in prose, as in the famous example of Monsieur Jourdain, so they can behave rationally without knowing the rules that determine rationality and without even being aware that they are following those rules. But we, as theorists, have as our aim to discover and formulate those rules.

3. *Rationality is a separate cognitive faculty.*

According to Aristotle, and a distinguished tradition that he initiated, the possession of rationality is our defining trait as humans: Man is a rational animal. Nowadays the fashionable term for faculty is "module," but the general idea is that humans have various special cognitive capacities, one for vision, one for language, etc. and rationality is one of these special faculties, perhaps even the most distinctive of our human capacities. A recent book even speculates on the evolutionary advantages of our having this faculty.[2]

4. *Apparent cases of weakness of will, what the Greeks called "akrasia," can only arise in cases where there is something wrong with the psychological antecedents of the action.*

Because rational actions are caused by beliefs and desires, and the beliefs and desires typically cause the action by first leading to the formation of an intention, apparent cases of weakness of will require a special explanation. How is it at all possible that an agent can have the right beliefs and desires, and form the right sort of intention, and still not perform the action? The standard account is that apparent cases of akrasia are all cases where the agent did not in fact have the right kind of antecedents to the action. Because the beliefs and desires, and derivatively the intentions, are causes, then if you stack them up rationally, the action will ensue by causal necessity. So in cases where the action does not ensue, there must have been something wrong with the causes.

Weakness of will has always been a problem for the Classical Model, and there is a lot of literature on the subject,[3] but weakness of will is always made out to be something very strange and hard to explain, something that could only happen under odd, or bizarre, circumstances. My own view, which I will explain later, is that akrasia is as common as wine in France. Anybody who has ever tried to stop smoking, lose weight or drink less at big parties will know what I am talking about.

5. *Practical reason has to start with an inventory of the agent's primary ends, including the agent's goals and fundamental desires, objectives, and purposes; and these are not themselves subject to rational constraints.*

In order to engage in the activity of practical reasoning, an agent must first have a set of things that he or she wants or values, and then practical reasoning is a matter of figuring out how best to satisfy this set of desires and values. We can state this point by saying that in order for practical reasoning to have any field in which to operate, the agent must begin

2. Nozick, Robert, *The Nature of Rationality*, Princeton, Princeton University Press, 1993. Chapter I.
3. For an anthology see *Weakness of Will*, edited by G. W. Mortimore, Macmillan St. Martin's Press, London, 1971.

with a set of primary desires, where desires are construed broadly, so that moral evaluations and sorts of things that the agent values, would count as desires. But unless you have some such set of desires to start with, there is no scope for reason, because reason is a matter of figuring out what else you ought to desire, given that you already desire something. And those primary desires are not themselves subject to rational constraints.

The model of practical reason is something like the following: Suppose you want to go to Paris, and you reason how best to go. You could take a ship or go by kayak or take an airplane, and finally after the exercise of practical reason, you decide to take the airplane. But if this is the only way that practical reason can operate, by figuring out "means" to "ends," two things follow: first, there can be no reasons for action which do not arise from desires, broadly construed. That is, there cannot be any desire-independent reasons for action. And second, those initial or primary desires cannot themselves be rationally evaluated. Reason is always about the means never about the ends.

This is at the heart of the Classical Model. When Hume said, "Reason is and ought only to be the slave of the passions" he is usually interpreted as making this claim; and the same claim made by many recent authors. For example, Herbert Simon writes, "Reason is wholly instrumental. It cannot tell us where to go; at best it can tell us how to get there. It is a gun for hire that can be employed in the service of any goals that we have, good or bad."[4] Bertrand Russell is even more succinct: "Reason has a perfectly clear and concise meaning. It signifies the choice of the right means to an end that you wish to achieve. It has nothing whatever to do with the choice of ends."[5]

6. *The whole system of rationality only works if the set of primary desires is consistent.*

A typical expression of this view is given by Jon Elster. "Beliefs and desires can hardly be reasons for action unless they are consistent. They must not involve logical, conceptual, or pragmatic contradictions."[6] It is easy to see why this seems plausible: if rationality is a matter of reasoning logically, there cannot be any inconsistencies or contradictions in the axioms. A contradiction implies anything, so if you had a contradiction in your initial set of desires, anything would follow, or so it seems.

III. Some Doubts About the Classical Model

I could continue this list, but even this much gives the general flavor of the concept; and I want in what follows to give some reasons why I think every one of these claims is false. At best they describe special cases, but they do not give a general theory of the role of rationality in thought and action.

1. *Rational actions are not caused by beliefs and desires. In general only irrational or non-rational actions are caused by beliefs and desires.*

4. *Reason in Human Affairs*, Stanford, CA: Stanford University Press, 1983, pp. 7–8.
5. *Human Society in Ethics and Politics*, London: Allen and Unwin, 1954, p. viii.
6. *Sour Grapes: Studies in the Subversion of Rationality* Cambridge:Cambridge University Press, 1983, p. 4.

Let us start, as an entering wedge, with the idea that rational actions are those that are caused by beliefs and desires. It is important to emphasize that the sense of "cause" is the ordinary "efficient cause" sense in which the explosion caused the building to collapse, or the earthquake caused the destruction of the freeway. I want to say that far from being the model of rationality, where the belief and the desire really are causally sufficient conditions of the action is in bizarre and typically irrational and non-rational cases. Those are the cases where, for example, the agent is in the grip of an obsession or an addiction. In a typical rational case where, for example, I am trying to decide which candidate to vote for, I consider various reasons for my decision. But I can only engage in this activity if I assume that the set of beliefs and desires by themselves are not causally sufficient to determine the action. The operation of rationality presupposes that there is a gap between the set of intentional states on the basis of which I make the decision, and the actual making of the decision. That is, unless I presuppose that there is a gap, I cannot get started with the process of rational decision making. To see this point you need only consider cases where there is no gap, where the belief and the desire are really causally sufficient. This is the case, for example, where the drug-addict has an overpowering urge to take heroin, he believes that this is heroin; so, compulsively, he takes it. In such a case the belief and the desire are sufficient to determine the action, because the addict cannot help himself. But that is hardly the model of rationality. Such cases seem outside the scope of rationality altogether.

In the normal case of rationality, we have to presuppose that the antecedent set of beliefs and desires is not causally sufficient to determine the action. This is a presupposition of the process of deliberation, and is absolutely inescapable for the application of rationality. We presuppose that there is a gap between the "causes" of the action in the form of beliefs and desires and the "effect" in the form of the action. This gap has a traditional name. It is called "the freedom of the will". In order to engage in rational decision making we have to presuppose free will. Indeed as we will see later, we have to presuppose free will in any rational activity whatever. We cannot avoid the presupposition, because even a refusal to engage in rational decision making is only intelligible to us as a refusal, if we take it as an exercise of freedom. To see this consider examples. Suppose you go into a restaurant, and the waiter brings you the menu. You have a choice between, let's say, veal chops and spaghetti; you cannot say: "Look, I am a determinist, Chè serà, serà. I will just wait and see what I order! I will wait to see what my beliefs and desires cause." This refusal to exercise your freedom, is itself only intelligible as an exercise of freedom. Kant pointed this out a long time ago: There is no way to think away your own freedom in the process of voluntary action because the process of deliberation itself can only go on against the presupposition of freedom, against the presupposition that there is a gap between the causes in the form of your beliefs, desires and other reasons, and the actual decision that you make.

If we are going to speak precisely about this, I think we must say that there are (at least) three gaps. First, there is the gap of rational decision making, where you try to make up your mind what you are going to do. Here the gap is between the reasons for making up your mind, and the actual decision that you make. Second, there is a gap between the decision and the action. Just as the reasons for the decision were not causally sufficient to produce the decision, so the decision is not causally sufficient to produce the action. There

comes the point, after you have made up your mind, when you actually have to do it. And once again, you cannot sit back and let the decision cause the action, any more than you can sit back and let the reasons cause the decision. For example, let us suppose you have made up your mind that you are going to vote for candidate Jones. You go into the voting booth with this decision firmly in mind, but once there you still have to do it. And sometimes, because of this second gap, you just do not do it. For a variety of possible reasons – or maybe none – you do not do the thing you have decided to do.

There is a third gap that arises for actions and activities extended in time, a gap between the initiation of the action and its continuation to completion. This gap exists for any temporally extended act, but it is most obvious in the case of complex actions. Suppose, for example, that you have decided to learn Portuguese, swim the English Channel or write a book about rationality. There is first the gap between the reasons for the decision and the decision, second the gap between the decision and the initiation of the action, and third there is a gap between starting the task and its continuation to completion. Even once you have started you cannot let the causes operate by themselves, you have to make a continuous voluntary effort to keep going with the action or activity to its completion.

At this point of the discussion I want to emphasize two points: the existence of the gap(s) and the centrality of the gap for the topic of rationality.

What is the argument for the existence of the gap(s)? The simplest arguments are the ones I just gave. Consider any situation of rational decision making and acting and you will see that you have a sense of alternative possibilities open to you and that your acting and deliberating make sense only on the presupposition of those alternative possibilities. Contrast these situations with those where you have no such sense of possibilities. In a situation in which you are in the grip of an overpowering rage, so that you are, as they say, totally out of control, you have no sense that you could be doing something else.

Another way to see the existence of the gap is to notice that in a decision making situation you often have several different reasons for performing an action, yet you act on one and not the others and you know without observation which one you acted on. This is a remarkable fact, and notice the curious locution we have for describing it: *you acted on* such and such a reason. Suppose for example that you had a whole bunch of reason both for and against voting for Clinton in the presidential election. You thought he would be a better president for the economy but worse for foreign policy. You like the fact that he went to your old college but don't like his personal style. In the end you vote for him because he went to your old college. The reasons did not operate on you. Rather you *chose* one reason and acted on that one. You made that reason effective by *acting on it*.

This is why, incidentally, the explanation of your action and the justification may not be the same. Suppose you are asked to justify voting for Clinton, you might do so by appealing to his superior management of the economy. But it may be the case that the actual reason you acted on was that he went to your old college in Oxford, and you thought, "College loyalty comes first." And the remarkable thing about this phenomenon is: in the normal case you know without observation which reason was effective, because you made it effective. That is to say, a reason for action is only an effective reason if you make it effective.

An understanding of the gap is essential for the topic of rationality because rationality can only operate in the gap. Though the concept of freedom and the concept of rationality

are quite different, the extension of rationality is exactly that of freedom. The simplest argument for this point is that rationality is only possible where irrationality is possible, and that requirement entails the possibility of choosing between various rational options as well as irrational options. The scope of that choice is the gap in question.

What fills the gap? Nothing. Nothing fills the gap: you make up your mind to do something, or you just haul off and do what you are going to do, or you carry out the decision you previously made, or you keep going, or fail to keep going, in some project that you have undertaken.

Even though we have all these experiences, could not the whole thing be an illusion? Yes it could. Our gappy experiences are not self validating. On the basis of what I have said so far, freedom could still be a massive illusion.

2. R*ationality is not entirely or even largely a matter of following rules of rationality.*

Let us turn to the second claim of the Classical Model, that rationality is a matter of rules, that we think and behave rationally to the extent that we think and act according to these rules. When asked to justify this claim, I think most traditional theorists would simply appeal to the rules of logic. And, an obvious kind of case that a defender of the Classical Model might present would be, let's say, a simple modus ponens argument.

If it rains tonight, the ground will be wet.
It will rain tonight.
Therefore, the ground will be wet.

Now, if you are asked to justify this inference, the temptation is to appeal to the rule of modus ponens: p, and if p then q, together imply q.

$$p \,\&\, (p \rightarrow q) \rightarrow q$$

But that is a fatal mistake. When you say that, you are in the Lewis Carroll Paradox.[7]

I will now remind you how it goes: Achilles and the tortoise are having an argument, and Achilles says (This is not his example but it makes the same point), "If it rains tonight, the ground will be wet, it will rain tonight, therefore the ground will be wet", and the tortoise says, "Fine, write that down, write all that stuff down", And when Achilles had done so he says, "I don't see how you get from the stuff before the "therefore" to the stuff after. What forces you to to make or even justifies you in making that move?" Achilles says, "Well that move rests on the rule of modus ponens, the rule that p, and if p then q, together imply q." "Fine." says the tortoise, "So write that down, write that down with all the rest" And when Achilles had done so the tortoise says, "Well we have all that written down, but I still don't see how you get to the conclusion, that the ground will wet." "Well don't you see?" says Achilles, "Whenever you have p, and if p then q, and you have the rule of modus ponens that says whenever you have p, and if p then q, you can infer q, then you can infer q." "Fine", says the tortoise, "now just write all that down." And you see

7. Carroll, Lewis "What Achilles Said to the *Tortoise*", *Mind*, 1895.

where this is going to go. We are off and running with an infinite regress.

The way to avoid an infinite regress is to refuse to make the first fatal move of supposing that the rule of modus ponens plays *any role whatever* in the validity of the inference. The derivation does not get its validity from the rule of modus ponens, rather the inference is perfectly valid as it stands without any outside help at all. It would be more accurate to say that the rule of modus ponens gets its validity from the fact that it expresses a pattern of an infinite number of inferences that are independently valid. The actual argument does not get its validity from any external source: if it is valid, it can only be valid because the premises entail the conclusion. Because the meanings of the words themselves are sufficient to guarantee the validity of the inference, we can formalize a pattern which describes an infinite number of such inferences. But the inference does not derive its validity from the pattern. The so-called "rule" of modus ponens is just a statement of a pattern of an infinite number of such independently valid inferences. Remember: *If you think that you need a rule to infer q from p and (if p then q), then you would also need a rule to infer p from p.*

What goes for this argument goes for any valid deductive argument. Logical validity does not derive from the rules of logic.

It is important to understand this point precisely. It is usually said that the mistake of Achilles was to treat modus ponens as another premise and not as a rule. But that is wrong. Even if he writes it down as a rule and not a premise, there would still be an infinite regress. It is equally wrong (indeed it is the same mistake) to say that the derivation derives it's validity from both the premises and the rule of inference.[8] The correct thing is to say that the rules of logic play no role whatever in the validity of valid inferences. The arguments, if valid, have to be valid as they stand.

We are actually blinded to this point by our very sophistication because the achievements of proof theory have been so great, and have had such important payoffs in fields like Computer Science, that we think that the syntactical analogue of *modus ponens* is really the same thing as the "rule" of logic. But they are quite different. If you have an actual rule that says whenever you see, or your computer "sees," a symbol with this shape

$$p$$

followed by one with this shape

$$p \rightarrow q$$

You or it writes down one with this shape

$$q$$

You have an actual rule that you can follow and that you can program into the machine so as to causally affect its operations. This is a proof-theoretical analogue of the rule of modus ponens, and it really is substantive, because the marks that this rule operates over are just meaningless symbols. The rule operates over otherwise uninterpreted formal elements.

8. For an example of this claim see Railton, Peter, "On the Hypothetical and the Non-Hypothetical in Reasoning about Belief and Action" pp. 53–79 in Cullity, G. and Gaut, B., *Ethics and Practical Reason*, Oxford: Oxford University Press, 1997, pp. 76–79.

Thus are we blinded to the fact that in real life reasoning, the rule of modus ponens plays no justificatory role at all. We can make proof theoretical or syntactical models, where the model exactly mirrors the substantive, or contentful processes of actual human reasoning. And of course, as we all know, you can do a lot with the models. If you get the syntax right, then you can plug in the semantics at the beginning and it will go along for a free ride, and you get the right semantics out at the end because you have the right syntactical transformations.

There are certain famous problems, most famously Gödel's Theorem, but if we leave them to one side, the sophistication of our simulations enables us to forget the actual semantic content when we make machine models of reasoning. But in real life reasoning the semantic content is what guarantees the validity of the inference, not the syntactical rule.

There are two important philosophical points to be made about the Lewis Carroll paradox. The first, that I have been belaboring, is that the rule plays no role whatever in the validity of the inference. The second is about the gap. *We need to distinguish between entailment and validity as logical relations on the one hand, and inferring as a voluntary human activity on the other.* In the case we considered, the premises entail the conclusion, so the inference is valid. But there is nothing that forces any actual human being to make that inference. You have the same gap for the human activity of inferring as you do for any other voluntary activity. Even if we convinced both Achilles and the Tortoise that the inference was valid as it stands and that the rule of modus ponens does not lend any validity to the inference, all the same, the tortoise might still, irrationally, refuse to make the inference. The gap applies even to logical inferences.

I am not saying that there could not be any rules to help us in rational decision making. On the contrary there are many famous such rules and even maxims. Here are some of them: "A stitch in time saves nine." "Look before you leap." "He who laughs last laughs best". And my favorite, "Le coeur a ses raisons que la raison ne connait pas." What I am saying is that rationality is not constituted as a set of rules, and rationality in thought as well as in action is not defined by any set of rules.

3. *There is no separate faculty of rationality.*

It should be implicit in what I have said that there cannot be a separate faculty of rationality distinct from such capacities as those for language, thought, perception and the various forms of intentionality, because rational constraints are already built into, they are internal to, the structure of intentionality in general and language in particular. Once you have intentional states, once you have beliefs and desires and hopes and fears, and, especially, once you have language, then you already have the constraints of rationality. That is, if you have a beast that has the capacity for forming beliefs on the basis of its perceptions, and has the capacity for forming desires in addition to beliefs, and also has the capacity to express all this in a language, then you have already have the constraints of rationality built into that structure. Rationality is not a separate faculty; it is built into the

structure of thought and language. To make this clear with an example: there is no way you can make a statement without caring about such questions such as, "Is it true or false?" "Is it consistent, or inconsistent with other things I have said?" So, the constraints of rationality are not an extra faculty that come in addition to intentionality and language. Once you have intentionality and language, you have already have phenomena that internally and constitutively have the constraints of rationality.

I like to think of it this way: The constraints of rationality ought to be thought of adverbially. They are a matter of the way in which we coordinate our intentionality. They are a matter of the way in which we coordinate the relations between our beliefs, desires, hopes, fears, and perceptions, and other intentional phenomena.

That coordination presupposes the existence of the gap. It presupposes that the phenomena at any given point are not causally sufficient to fix the rational solution to a problem. And I think we can now see why the same point operates for theoretical as for practical reason. If I hold up my hand in front of my face, there is no gap involved in seeing my hand, because I cannot help seeing my hand in front of my face. It is not up to me. So there is no question of such a perception being either rational or irrational. But now, suppose I refuse to believe that there is a hand in front of my face, even in this situation where I cannot help seeing it. Suppose I just refuse to accept it: "You say there's a hand there but I damn well refuse to accept that claim" Now the question of rationality arises, and I think we would say that I am being irrational in such a situation.

I want to emphasize a point I made earlier. You can only have rationality where you have the possibility of irrationality. And with just sheer, raw perceptions, you do not get rationality or irrationality. They only come into play where you have a gap, where the existence of the intentional phenomena by themselves is not sufficient to cause the outcome, and these are cases where you have to decide what you are going to do or think.

This is why people whose behavior is determined by sufficient causal conditions are removed from the scope of rational assessment. For example, not long ago I was in a committee meeting, and a person whom I had previously respected voted in the stupidest possible way. I said to him afterwards, "How could you have voted that way on that issue?" And he said, "Well, I'm just incurably politically correct. I just can't help myself." His claim amounts to saying that his decision making in this case was outside the scope of rational assessment, because the apparent irrationality was a result of the fact that he had no choice at all, that the causes were causally sufficient.

4. *Weakness of will is a common, natural form of irrationality. It is a natural consequence of the gap.*

On the Classical Model, cases of weakness of will are strictly speaking impossible. If the antecedents of the action are both rational and causal, and the causes set sufficient conditions then the action has to ensue. It follows, that if you did not do the thing you set out to do, then that can only be because there was something wrong with the way you set up the antecedents of the action. Your intention was not the right kind of intention,[9] or you were not fully morally committed to the course you claimed to be committed to.[10]

I want to say, on the contrary, that no matter how perfectly you structure the antecedents of your action, weakness of will is always possible. Here is how: at any given point in

our waking lives, we are confronted with an indefinitely large range of possibilities. I can raise my right arm, or I can raise my left arm, I can put my hat on top of my head, or I can wave it around. I can drink water or not drink water. More radically, I can walk out of the room and go to Timbuktu, or join a monastery, or do any number of other things. I have an open-ended sense of possibilities. Now, of course, in real life there will be restrictions set by my Background, by my biological limitations and by the culture that I have been brought up in. The Background restricts my sense of the possibilities that are open to me at any given time. I cannot, for example, in real life, imagine doing what St. Simeon Stylites did. He spent thirty five years on top of a pillar, just sitting there on a tiny platform, all for the glory of God. That is not an option that I could seriously consider. But I still have an indefinite range of real options that I am capable of perceiving as options. Weakness of will arises simply from the fact that at any point the gap provides an indefinitely large range of choices open to me and some of them will seem attractive even if I have already made up my mind to refuse them. It does not matter how you structure the causes of the action in the form of antecedent intentional states - beliefs, desires, choices, decisions, intentions – in the case of voluntary actions, the causes still do not set sufficient conditions and this opens the way for weakness of will.

It is an unfortunate feature of our philosophical tradition that we make weakness of will out to be something really strange, really bizarre; whereas, I have to say I think it is very common in real life.

5. *Contrary to the Classical Model there are desire independent reasons for action.*

The fifth thesis of the Classical Model that I want to challenge has a very long history in our philosophical tradition. The idea is this: A rational act can only be motivated by a desire, where "desire" is construed broadly to include moral values that one has accepted, and various sorts of evaluations that one has made. Desires need not be all egotistical, but for any rational process of deliberation there must be some desire that the agent had prior to the process, otherwise there would be nothing to reason from. There would not be any basis on which you could do your reasoning, if you did not have a set of desires in advance. Thus there can be no reasoning about ends, only about means. A sophisticated contemporary version of this view is in the work of Bernard Williams,[11] who claims that there cannot be any "external" reasons for an agent to act. Any reason which is a reason for the agent must appeal to something which is "internal" to his "motivational set." This amounts in my terminology to saying that there cannot be any desire-independent reasons for action. There is a great deal to be said about the problems raised by this view, but for present purposes I want to note only that it has the following absurd consequence: At any given point in one's life no matter what the facts are, and no matter what one has done in the past or knows about one's future, no one can have any reason to do anything unless right then and there, there is an element of the motivational set, a desire broadly con-

9. Donald Davidson, "How is Weakness of the Will Possible?" *Essays on Actions and Events*, Clarendon Press, Oxford, Oxford University Press, New York, 1980.
10. R. M. Hare, *The Language of Morals*, Oxford: Oxford University Press, 1952.
11. "External and Internal Reasons" reprinted in his *Moral Luck: Philosophical Papers 1973–1980*, Cambridge: Cambridge University Press, 1981, pp. 101–13.

strued, to do that thing or a desire for which doing that thing would be a "means" to that "end", that is, a means to satisfying that desire.

Now why is that absurd? Well, try to apply it to real life examples. Suppose you go into a bar and order a beer. They bring the beer and you drink it. They bring you the bill and you say to them. "I have looked at my motivational set and I find no internal reason for paying for this beer. None at all. Ordering and drinking the beer is one thing, finding something in my motivational set is something else. The two are logically independent. Paying for the beer is not something I desire for its own sake, nor is it a means to an end or constitutive of some end that is represented in my motivational set. I have read Professor Williams, and I have also read Hume on this subject, and I looked carefully at my motivational set, and I cannot find any desire there to pay this bill! I just can't! And therefore, according to all the standard accounts of reasoning, I have no reason whatever to pay for this beer. It is not just that I don't have a strong enough reason, or that I have other conflicting reasons, but I have zero reason. I looked at my motivational set, I went through the entire inventory, and I found no desire, either primary or derived, to pay for the beer."

We find this speech absurd because we understand that when you ordered the beer and drank it, if you are a sane and rational person, you were intentionally *creating* a desire independent reason, a reason for doing something regardless of what was in your motivational set when the time came to do it. The absurdity lies in the fact that on the Classical Model the existence of a reason for an agent to act depends on the existence of a certain sort of psychological element in his motivational set, it depends on the existence of a desire, broadly construed, then and there; and in the absence of that desire the agent has no reason, regardless of all the other facts about him and his history, and regardless of what he knows. But in real life the sheer knowledge of external facts in the world, such as the fact that you ordered the beer and drank it, can be a rationally compelling reason to pay for it.

There are really two strands to this aspect of the Classical Model. First we are supposed to think that all reasoning is about means not about ends, that there are no external reasons for action. And secondly, a corollary that the primary ends in the motivational set are outside the scope of reason. Remember that Hume also says, "Tis not contrary to reason to prefer the destruction of the whole world to the scratching of my finger".[12] The way to assess any such claim is always to bring it down to real life cases. Suppose the president of the United States went on television and said, "I have consulted with the Cabinet and the leaders of Congress, and I have decided that there's no reason why I should prefer the scratching of my little finger to the destruction of the whole world". If he did this in real life we would feel he had, to use the terminology of Hume's era, "lost his reason". There is something fishy about Hume's claim and the general thesis that one's fundamental ends can be anything whatever, and are totally outside the scope of rationality, that where primary desires are concerned, everything has equal status and is equally arbitrary. I think that cannot be the right way to look at these matters.

The thesis that there are no desire-independent reasons for action, that there are no external reasons, is logically closely related to Hume's doctrine that one cannot derive an "ought" from an "is". Here is the connection. "Ought" statements express reasons for action. To say that someone ought to do something is to imply that there is a reason for him to do it. So Hume's claim amounts to the claim that statements asserting the existence of reasons for action cannot be derived from statements about how things are. But how things are is a matter of how things are in the world as it exists independent of the agent's motivational set. So on this interpretation, the claim that how things are in the world cannot imply the existence of any reasons in an agent's motivational set (one cannot derive "ought" from "is") is closely related to the claim that there are no facts in the world independent of the agent, that by themselves constitute reasons for action (there are no external reasons). Hume says, in effect, we cannot get values from facts, Williams says that cannot get motivations from external facts by themselves. The point of connection lies in the fact that the acceptance of a value is the acceptance of a motivation. However we interpret both claims, I think they are both demonstrably false. The simplest form of demonstration is to show how we can create rationally binding desire-independent reasons for action. It is a non-trivial question how we do this, and I do not have the time or space here to develop it in detail. I go into it in considerably more detail in more forthcoming book on rationality. But the basic idea is that rational agents, in speaking a language, can voluntarily and intentionally create desire-independent reasons for themselves to perform some course of action in the future. The most obvious cases of this are, of course, promises, but the element of commitment pervades almost all speech acts and each such commitment is a desire-independent reason for action.

12. Hume, *Treatise of Human Nature*, Oxford, 1888, p. 416.

6. *Inconsistent reasons for action are common and indeed inevitable. There is no rational requirement that rational decision making must start with a consistent set of desires or other primary reasons for acting.*

The last point I want to take up is the question of consistency. As with the argument about weakness of will, this part of the Classical Model – the claim that the set of primary desires from which one reasons must be consistent – does not seem to me just a little bit false, but radically mistaken. It seems to me that most practical reasoning is typically about adjudicating between conflicting, inconsistent desires and other sorts of reasons. Right now, today, I very much want to be in Paris but I also want very much to be in Berkeley. And this is not a bizarre situation, rather it seems to me typical that we have an inconsistent set of ends. Given the extra premise that I know I cannot be both in Berkeley and in Paris at the same time, I have a logically inconsistent set of desires, and the task of rationality, the task of practical reason, is to try to find some way to adjudicate between these various inconsistent aims. Typically in practical reasoning you have to figure out how to give up on satisfying some desires in order to satisfy others. The standard way out of this problem in the literature is to say that rationality is not about desires as such but about *preferences*. Rational deliberation must begin with a well ordered preference schedule. The problem with that answer is that in real life deliberation is largely about forming a set of preferences. A well ordered set of preferences is typical the *result* of successful deliberation, and is not its *precondition*. Which do I prefer, to be in Berkeley or Paris? Well, I would have to think about it.

And even after you have made up your mind, you decide "O.K., I'm going to Paris", that decision itself introduces all sorts of other conflicts. You want to go to Paris, but you

do not want to stand in line at airports, you do not want to eat airplane food, you do not want to sit next to people who are trying to put their elbow where you are trying to put your elbow. And so on. There are just all kinds of things that you do not want to happen, which you know are going to happen once you try to carry out your decision to go to Paris and to go by plane. The point I want to emphasize is that there is a long tradition associated with the Classical Model, whereby inconsistent reasons for action, such as inconsistent obligations, are supposed to be philosophically odd or unusual. Often people in the tradition try to fudge the inconsistencies by saying that some of the apparently inconsistent obligations are not real honest-to-john obligations, but mere "prima facie" obligations. But rational decision making is typically about resolving between conflicting reasons for action and you only have a genuine conflict of obligations where they are all genuine obligations.

IV. Conclusion

This talk is really only the beginning of what has to be a much larger investigation. The real subject matter of rationality is human freedom, for reasons that I have hinted at in this brief discussion. The very attempt to exercise rationality, and indeed the possibility of irrationality, presuppose the gap. Rationality is typically about resolving conflicting reasons for action in the gap; and it is only because of the gap that we have the possibility of freely creating desire independent reasons for action. These features make the subject of rationality much more difficult than is envisioned by the Classical Model, but at the same time they make it much more interesting.

Wann ist die Vernunft praktisch und wann Normativität moralisch?

ULRICH STEINVORTH

Zu den Fragen

Was haben Vernunft und Moral miteinander zu tun? Unbestritten kann es ohne Vernunft keine Moral geben. Aber ist Moralischsein *dasselbe* wie Vernünftigsein? Hume verneinte: „Reason is, and ought only to be the slave of the passions"[1]; Kant hielt dagegen, „daß es *reine* praktische Vernunft gebe"[2], und implizierte, daß unser Handeln genau dann nur von der Vernunft bestimmt wird, wenn wir moralisch handeln. Tatsächlich ist zwar auch nach Hume die Vernunft praktisch, jedoch nur dann, wenn sie Mittel für Zwecke angibt, die nie die Vernunft, sondern die *passions* setzen. Daher meine erste Frage: *Wann* ist die Vernunft praktisch? Meine zweite Frage ist durch Christine Korsgaard angeregt, die Kants These verteidigt. Sie reformuliert Kant durch die These, die Normativität der Moral sei die Normativität des Rationalen. Daher frage ich zweitens: Wann ist Normativität *moralisch*?

Zum Begriff der Moral

Was sollen wir unter *Moral* verstehen? Das Verständnis der Moralphilosophen von ihrem Gegenstand geht weit auseinander. Aber sie haben den Begriff der Moral nicht erfunden, sondern an bekannten Phänomenen orientiert. Alle Menschen werden zu Einstellungen und Handlungsweisen erzogen, die sie als moralisch vom Rest unterscheiden. In allen Gesellschaften gibt es einen Komplex von Regeln und Gesichtspunkten zur Bewertung von Handlungsweisen und Einstellungen. In Anlehnung an Sidgwick nenne ich ihn *ordinäre Moral*.[3] Sie ist nicht in allen Gesellschaften dieselbe und gewöhnlich in einer Gesellschaft weder konsistent noch ausreichend, für jede mögliche Handlung zu entscheiden, ob sie richtig ist. Aber sie ist der Gegenstand der Analyse und oft genug der Kritik und Reform von Moraltheoretikern. Gäbe es keine ordinäre Moral, so könnte es keine philosophische oder reflektierte Moral geben.

Muß es *inhaltliche* Gemeinsamkeiten zwischen den ordinären Moralen verschiedener Gesellschaften geben? Angesichts der kulturellen Unterschiede zwischen menschlichen

1. David Hume, A Treatise of Human Nature, ed. Selby-Bigge, Oxford 1978, 415.
2. Immanuel Kant, Kritik der praktischen Vernunft, ed. Vorländer, Hamburg 1959, 3. Meine Hervorhebung.
3. Henry Sidgwick, The Methods of Ethics, Indianapolis 1981 (Nachdr. 7. Aufl. 1907), spricht gewöhnlich von der *Morality of Common Sense*, vgl. Index des Buchs, aber auch von *ordinary morality* (36) und *ordinary practical thought* (6).

B. Brogaard, B. Smith (Hg.), *Rationality and Irrationality / Rationalität und Irrationalität*, S. 325–340.

Gesellschaften verneinen das manche Theoretiker. Regeln, die etwa das willkürliche Quälen von Kindern erlauben, wären demnach moralisch, wenn sie nur eine gewisse gesellschaftliche Verbreitung haben. Mir scheint diese Konsequenz zu sehr unseren Intuitionen von dem, was moralisch ist, zu widersprechen. Wenn eine Gesellschaft Bewertungsregeln folgt, die Grausamkeit, Willkür und Zerstörung zum Ideal erheben, ist es falsch, sie moralisch zu nennen.

In jedem Fall sollte die Frage, ob Vernünftigsein und Moralischsein dasselbe sind, klären, in welchem Verhältnis zur Vernunft nicht ein normatives System beliebigen Inhalts steht, sondern eines, das inhaltliche Mindestforderungen stellt, etwa nach folgender Definition[4]:

> Moral =df ein normatives System, das Wohltun unter Freunden gebietet, Achtung auch vor Fremden empfiehlt und Destruktivität auch gegen Feinde verbietet.

Diese Definition bestimmt zwar nur die *ordinäre* Moral. Diese könnte sich bestimmten Vernunftansprüchen nicht gewachsen zeigen. Dann wäre eine *reflektierte* Moral notwendig, die anders zu definieren wäre. Aber auch deren normativer Inhalt müßte aus dem der ordinären Moral entwickelt werden können.

Eine Reformulierung von Humes und Kants Thesen zur Moral

Was sollen wir unter *Vernunft* verstehen? „All parties can agree", sagt Robert Brandom, „that to be rational is to distinguish good inferences from bad inferences." Die Frage ist nur, wie er hinzufügt, „whether ‚good inference' in this formula can be restricted to *logically* good inferences, or again to *instrumentally* good inferences"[5], oder ob zu ihnen auch materiale Schlüsse gehören, wie Brandom meint[6], oder ob man die Vernunft sogar mit Joseph Raz als „the ability to realize the normative significance of the normative features of the world, and the ability to respond accordingly" definieren soll[7]. Je nachdem zu welcher Definition man kommt, präjudiziert man, ob Moralisch- und Vernünftigsein dasselbe sind. Raz' Definition entscheidet für Kant; die Definition der Vernunft über das logische und das instrumentelle Folgern für Hume, und je nachdem, wie weit man materiale Schlüsse faßt, folgt man Hume oder Kant.

4. Vgl. U. Steinvorth, Gleiche Freiheit, Berlin: Akademie 1999, 48–52 und ders., Ist Moral Zwang gegen sich selbst? In Brigitte Boothe, Hg., Moral – Gift oder Gottesgabe, Göttingen (Vandenhoeck) 2001.
5. Roger Brandom, Making It Explicit, Harvard UP 1994, 231. David Gauthier, Individual Reason, in J.B. Schneewind, Reason, Ethics, and Society, Chicago: Open Court, 1996, 39–57, 57f, unterscheidet *reason* als *capacity to act for reasons* und *rationality* als *capacity for assessing reasons*. Ich unterscheide nicht zwischen Vernunft und Rationalität.
6. Brandom ebd. 98, 134ff, 168. ‚A ist rechts von B, also ist B ist links von A'; ‚Es donnert, also hat es geblitzt' sind materiale Schlüsse.
7. Joseph Raz, Explaining Normativity: On Rationality and the Justification of Reason, in Engaging Reason. On the Theory of Value and Action, Oxford 1999, 67–89; 69 und 68. Ähnlich Nicholas Rescher, Rationalität. Eine philosophische Untersuchung über das Wesen und die Rechtfertigung von Vernunft, Würzburg (Königshausen & Neumann) 93, Kap. 6, 109ff.

Wir könnten daher versuchen, den Streit zwischen Hume und Kant daran zu entscheiden, wie man das Folgern definieren muß. Ich halte zwar daran fest, daß die Vernunft ein Vermögen ist, das mit der Beurteilung von Gründen zu tun hat, aber gehe nicht der Frage nach, was richtiges Folgern ist oder wie man Gründe zu beurteilen hat. Ich möchte lieber im Auge behalten, worin Hume und Kant in ihrer Sicht des Verhältnisses von Vernunft und Moral auseinandergehen.

Hume und Kant sind darin einig, *daß* wir für moralische Entscheidungen *Gründe* haben. Sie streiten darüber, welcher Art die Gründe sind. Kant behauptete, sie seien *Gründe für das Handeln überhaupt*. Hume hielt sie für *moralspezifische* Gründe. Er nahm für moralische Urteile eine *besondere* Handlungsrationalität an, die mit den *passions* zu tun hat, deren Sklavin die Vernunft sei; Kant dagegen eine *allgemeine*. Ich schlage daher folgende Formulierung für Humes und Kants Thesen vor. Sie unterstellt nicht, daß Hume und Kant die in ihr vorkommenden Begriffe gebrauchen. Daß diese dennoch sinnvoll sind, soll sich im folgenden zeigen.

> *Humes These*:
> Wenn wir moralisch handeln, folgen wir Prinzipien einer *besonderen Rationalität*. Die Gründe für moralisches Handeln unterscheiden moralisches von anderem Handeln.

> *Kants These*
> Wenn wir moralisch handeln, folgen wir nur Prinzipien der *allgemeinen Rationalität*. Die Gründe für moralisches Handeln sind Gründe für das Handeln überhaupt.

Diese Formulierung erlaubt uns ohne Präjudiz für eine Definition des Folgerns zu fragen, welcher Art die Rationalität moralischer Entscheidungen ist. Sie erlaubt uns auch, Korsgaards Kantverteidigung zu prüfen. Denn der Vernunftbegriff, den sie entwickelt und Kant, wie ich denke, zu Recht, unterstellt, ist der der allgemeinen Rationalität.

Die allgemeine Rationalität

Korsgaard beschreibt die allgemeine Rationalität, wenn sie das Problem des „self-conscious" Geistes beschreibt[8]. Ich dehne ihre Beschreibung hier etwas aus. Auch das tierische Hirn verarbeitet Reize, die mit einer Meinung oder einer Handlung enden können, etwa der Meinung, daß auf dem Tisch eine saftige Birne liegt, oder der Handlung, die Birne zu ergreifen und zu essen. Solche Reizverarbeitungen können hohe Intelligenz bezeugen, wenn etwa die Birne schwer zu erkennen oder zu erreichen ist. Beim Menschen oder dem selbstbewußten Geist aber werden die spontanen Reizverarbeitungen reflektiert und gebrochen, nämlich wenn wir fragen, ob die Meinung nicht täuscht oder die Absicht nicht enttäuschen wird; ob das, was wir für eine Birne halten, wirklich eine Birne ist, oder ob der Verzehr der Birne uns gut tut. Beim Menschen treten die „gaps" auf, die Searle hier immer wieder als Eigenart menschlicher Reizverarbeitung hervorgehoben hat.[9] Die Lö-

8. Christine M. Korsgaard, The Sources of Normativity, Cambridge UP 1996, 92f.
9. In seinen Veröffentlichungen etwa in John Searle, Mind, Language and Society. Doing Philoso-

sung des Problems des „selbstbewußten Geistes" besteht darin, daß wir nach *Gründen* für die spontanen Meinungen und Handlungen oder ihre Alternativen suchen. „We need reasons because our impulses must be able to withstand reflective scrutiny."[10] Die allgemeine Rationalität ist das, was wir betätigen, wenn wir Fragen stellen und beantworten, die bei Brechung der spontanen Reizverarbeitung auftreten.

Was ich *allgemeine* Rationalität nenne, nennt Korsgaard einfach *reason*. In der Tat hat das, was wir in der gebrochenen Reizverarbeitung betätigen, Qualitäten, die man der Vernunft traditionell zuschrieb, ohne zwischen allgemeiner und besonderer Rationalität zu unterscheiden Korsgaard zählt sie nicht auf, aber wir sollten sie hervorheben. Es sind folgende Qualitäten. Die Betätigung zeichnet erstens Menschen vor Tieren aus; die Vernunft galt traditionell als die Auszeichnung des Menschen. Sie erschüttert zweitens vermeintliche Gewißheiten. Wegen dieser Qualität wurden sokratische Fragen zum Muster der Vernunfttätigkeit. Drittens, obgleich sie Gewißheiten erschüttert, entscheidet nur sie über Richtigkeitsansprüche. Solche Ansprüche werden erst bei gebrochener Reizverarbeitung möglich, aber auch nötig, da die Unterscheidung von *richtig* und *falsch* bei ungebrochener Reizverarbeitung keinen Platz hat. Viertens entscheidet sie durch *Gründe*. Die Reize, denen das Tier ungebrochen folgt, werden nach Brechung ihrer Spontaneität auf ihre Vertrauenswürdigkeit betrachtet und dadurch als Gründe behandelt.[11]

Die gebrochene Reizverarbeitung hat auch Folgen für die Natur des Lebewesens, die ebenfalls den traditionellen Vernunftbegriff bestimmten. Wenn wir Reize auf ihre Vertrauenswürdigkeit prüfen und sie als Gründe behandeln, genügt es nicht, nur die vom Reiz spontan ausgelöste Meinung oder Handlung und ihre Folgen zu betrachten. Wir müssen sie vielmehr mit Alternativen vergleichen. Unsere Entscheidung ist um so begründeter, je mehr Alternativen wir erwägen. *Vollkommen* ist sie nur dann begründet, wenn wir *alle* Alternativen betrachtet haben. Wir können aber zu keiner Zeit sicher sein, alle möglichen Alternativen betrachtet zu haben, weil immer neue auftauchen können. Als letzter Richter über alle Zweifel verlangt die Vernunft möglichste *Vollständigkeit* der Entscheidungsgründe und nimmt dadurch die merkwürdige Eigenschaft an, uns ohne Rücksicht auf einen unmittelbaren Nutzen zur Erforschung möglichst aller Wirklichkeitsbereiche anzutreiben. Weil sie Richtigkeitsfragen entscheidet und Vollständigkeit des Materials verlangt, auf das sie ihre Entscheidungen stützt, gilt sie als *unbedingt* oder *absolut*. Manchen Philosophen galt sie auch als *unfehlbar*. Aber unfehlbar könnte die Vernunft nur dann sein, wenn sie der Vollständigkeit der möglichen Entscheidungsgründe sicher sein könnte. Das aber kann sie wegen des Wechsels der Reize nie. Sie fordert Vollständigkeit der Entscheidungsgründe, ist letztinstanzlich und fehlbar.

Die Brechung der spontanen Reizverarbeitung betrifft unsere Natur in einer vielleicht noch tiefer reichenden Weise. Sie macht unsere Meinungen und Handlungen nicht unbedingt intelligenter oder effektiver. Sie macht uns zu Subjekten, die ihre Reize nicht nur als potentielle Handlungsgründe, sondern auch als Objekte einer meinungsunabhängigen Welt behandeln. Ebenso notwendig macht uns die Brechung der spontanen Reizverarbeitung zu Wesen, die *normativ* sind. Dieser Punkt steht im Mittelpunkt von Korsgaards Interesse und wird von ihr anders als die bisher aufgezählten hervorgehoben.

phy in the Real World, London: Weidenfeld 1999, 107.

10. Ebd. 93.
11. Vgl. hierzu und zum folgenden U. Steinvorth, Warum überhaupt etwas ist, Reinbek b. Hamburg 1994, 20–104.

In seiner ungebrochenen Meinungs- und Absichtsbildung wird der Mensch vom Zwang der Reize bestimmt. Diese setzen ihm Ziele und legen ihn auf Mittel fest sie zu erreichen. In seiner *gebrochenen* Meinungs- und Absichtsbildung wird er vom Zwang der besseren Gründe bestimmt. Diese erlauben ihm, reizgesetzte Ziele und Mittel zu verwerfen und durch solche zu ersetzen, für die die besten Gründe sprechen. Folgt er den Gründen, so folgt er *auch* einer Art Zwang. Doch ist es nicht der von Reizen, sondern der der Vernunft. Er kann sich ihm widersetzen, aber nur bei Strafe der Unvernünftigkeit. Habermas beschreibt ihn als „eigentümlich zwanglosen Zwang des besseren Arguments"[12]. Korsgaard nennt ihn den Zwang der Normativität[13], und ich folge ihrem Gebrauch.

Nach Kant und Korsgaard, aber auch nach Habermas und Apel finden wir im Zwang der Normativität schon den der *Moral*. Tatsächlich aber haben wir soweit nur die Normativität der *allgemeinen* Rationalität beschrieben. Sie ergibt sich daraus, daß wir Normen brauchen, sobald unsere spontanen Handlungen gebrochen sind. Diese allgemeine Normativität ist unsere *Angewiesenheit* darauf, für Meinungen und Handlungen beliebiger Art auf *Gründe* zurückgreifen zu müssen. Sie ist das *Sollen*, auf das uns *beliebige* Gründe verweisen.

Die besondere Rationalität oder: Konstruktivität und Konstitutivität

Die besondere Rationalität ist die Normativität besonderer Gründe, die nicht unser Handeln überhaupt leiten, sondern unser Handeln in *besonderen* Materien. Korsgaard hat sie im Auge, wenn sie feststellt: „Concepts like knowledge, beauty, and meaning, as well as virtue and justice, all have a normative dimension, for they tell us what to think, what to like, what to say, what to do, and what to be. And it is the force of these normative claims – the right of these concepts to give laws to us – that we want to understand."[14]

Sie unterscheidet allerdings nicht zwischen allgemeiner und besonderer Normativität, offenbar weil sie die Unterscheidung für irrelevant hält. Tatsächlich ist sie für sie wie auch für Kant wichtig. Denn sie zeichnet unter den Begriffen, die eine normative Dimension haben, die der Tugend und Gerechtigkeit aus. Diese sagen uns, was wir *sein* sollen – die andern dagegen sagen uns, was wir meinen, lieben und sagen sollen. Korsgaard folgt in ihrer Auszeichnung der moralischen Regeln Kant. Kant versteht die Befolgung moralischer Regeln als die Bedingung der Möglichkeit, überhaupt verantwortlich und zurechenbar zu handeln, so wie er die Befolgung bestimmter Grundsätze des Verstandes als Bedin-

12. Jürgen Habermas, Erkenntnis und Interesse, Frankfurt 1973, 226 und 240; Theorie des kommunikativen Handelns, Frankfurt 1981, Bd.2, 52f und 48.
13. Christine Korsgaard, The Sources of Normativity, Cambridge UP 1996, spricht trotz des Buchtitels mehr von der *normative question* als von *normativity;* p. 226 scheint mir aber klarzumachen, daß sie *normativity* im Sinn von *rational necessity* versteht. Diese Notwendigkeit kann, wie Korsgaard ebd. 20f hervorhebt, verschiedener Art sein: „Obligation, the most obtrusively normative (concept), seems sternly to command; while beauty only to attract and meaning perhaps to suggest".
14. Korsgaard, The Sources of Normativity a.a.O. 9.

gung der Möglichkeit versteht, überhaupt Erfahrung und eine objektive Welt zu gewinnen. Moralische Regeln sind daher für Kant ebenso wie bestimmte Verstandesprinzipien *konstitutiv*. Sie konstituieren das Subjekt des Handelns; reine Verstandesprinzipien das Subjekt der Erfahrung und die Welt, die es erkennt und verändert.

Die Regeln dagegen, auf die Korsgaard mit den „concepts like knowledge, beauty, and meaning" verweist, sind nicht konstitutiv. Es sind Regeln der Prüfung wissenschaftlicher Theorien, ästhetische und semantische Regeln. Ihre Befolgung betätigt nicht die bisher beschriebene allgemeine Rationalität. Die Befolgung *konstitutiver* Regeln konstituiert ein Verhalten, das gründegeleitete Urteilen und Handeln eines autonomen oder verantwortlichen Subjekts, das das vorangehende reizgesteuerte Verhalten *vernichtet*. Sie beseitigt es zwar nicht vollständig aus dem menschlichen Leben, weil wir in unseren animalischen Funktionen reizgesteuert bleiben, vertreibt es aber nach dem Anspruch der konstitutiven Regeln restlos aus dem durch diese geschaffenen Reich des autonomen gründegeleiteten Entscheidens. Die Befolgung semantischer Regeln für den Gebrauch verbaler Empfindungsäußerungen beseitigt dagegen nicht die spontanen nichtverbalen Äußerungen, die sie regeln, sondern *entwickelt* sie. Die Ersetzung eines Schmerzensschreis durch eine verbale Äußerung erlaubt differenziertere Reaktionen auf den Schmerz als der Schrei: eine genauere Beschreibung und die Thematisierung von Schmerzen in Erörterungen verschiedenster Art. Sie macht trotzdem spontane und nichtverbale Schmerzäußerungen nicht überflüssig. In manchen Umständen bleibt man unfähig, Schmerzen anders als durch Schreien zu äußern, und ohne nichtverbale Äußerungen wäre das Erlernen verbaler Äußerungen gar nicht möglich.

Ähnlich legen Regeln der Prüfung wissenschaftlicher Theorien fest, wie spontane Annahmen über die Umwelt durch reflektierte zu ersetzen sind. Die neuen Annahmen, die die spontanen ersetzen, erlauben eine differenziertere Beschreibung der Welt, die über den Rahmen des Beobachtbaren und sogar des Vorstellbaren hinausgeht. Aber sie machen die spontanen nicht überflüssig; sie bleiben vielmehr zu ihrer Überprüfung immer auf die spontanen angewiesen. Auch diese sind *Beschreibungen*; nur keine wissenschaftlichen.

Ästhetische Regeln, so könnte man vielleicht sagen, schreiben vor, wie man wahrnehmbare Formen den Sinnen präsentieren soll, wenn man eine spezifisch ästhetische Zustimmung zu ihnen erreichen will. Eine solche Zustimmung könnte man damit umschreiben, daß sie weder von einem praktischen Nutzen noch einem begrifflichen Erkenntniszuwachs abhängt und dem Gegenstand der Zustimmung selbst gilt. Ästhetische Regeln ermöglichen Kunstwerke, die die spezifisch ästhetische Zustimmung differenzieren und auf Bereiche des alltäglichen und außeralltäglichen Lebens beziehen. Dadurch eröffnen sie ebenso wie die semantischen und die Regeln der Überprüfung wissenschaftlicher Theorien neue Dimensionen der Wirklichkeit. Aber ebensowenig wie sie machen sie die spontanen Reaktionen überflüssig. Denn auch diese sind *ästhetische* Reaktionen.

Die besondere Normativität der beschriebenen Regeln ist nicht *konstitutiv*, sondern *konstruktiv* in folgendem Sinn. Sie ersetzt die spontanen Handlungen nicht nur, sondern *entwickelt* ihre Anlagen – die Kommunikativität von Äußerungen, die Deskriptivität von Meinungen, die Sinnfälligkeit von Wahrnehmungen. Sie entwickelt sie so, daß sie eine Wirklichkeitsdimension, die in den spontanen Reaktionen angelegt ist, entfaltet. Die allgemeine Normativität konstitutiver Regeln sichert dagegen den Bestand eines Wesens,

das autonom handeln und sich die Welt in ihren verschiedenen Dimensionen und Möglichkeiten zugänglich machen kann.

Der Unterschied zwischen der Konstitutivität und der Konstruktivität von Regeln kennzeichnet den zwischen der allgemeinen und der besonderen Rationalität und Normativität. Wir brauchen hier nicht zu entscheiden, ob die allgemeine Rationalität unabhängig von der besonderen auftreten kann. Es genügt für unsere Frage anzuerkennen, daß Kant konstitutive Regeln annahm. Ein vielleicht besserer Kandidat für konstitutive Regeln als die moralischen und transzendentalen, die Kant annahm, sind die Regeln der Logik und der Argumentation. Sie ersetzen nichts, was neben ihnen fortbesteht. Sie ersetzen, wenn man so will, unlogisches Denken. Aber unlogisches Denken ist eben kein Denken, während der Schrei, den semantische Regeln durch eine verbale Äußerung zu ersetzen erlauben, selbst eine Äußerung ist. Auf der andern Seite beweist die Existenz logischer Regeln nicht, daß wir konstitutiven Regeln *allein* folgen können. Denn wenn wir logischen Regeln folgen, folgen wir offenbar zugleich auch semantischen, instrumentellen oder anderen nicht-logischen Regeln.

Ob wir aber konstitutiven Regeln allein folgen können oder nicht, es ist in jedem Fall sinnvoll, sie von konstruktiven Regeln zu unterscheiden. Denn mit ihnen können wir zwei Arten unterscheiden, auf Reize *reflektiert* zu reagieren. Man könnte sie *verwerfend* und *ausbauend* nennen. Folgen wir konstitutiven Regeln, so verwerfen wir die Reize, die uns bisher gesteuert und geformt haben, vollständig als Instanz der Steuerung unsres Verhaltens. Dies radikal negative Verhältnis zu Reizen (und Neigungen), für das Kant bekannt ist, ist notwendig, wenn es das autonome Subjekt geben soll, für dessen Annahme Kant ebenso bekannt ist. Folgen wir konstruktiven Regeln, so halten wir an der Steuerung durch Reize fest, aber beuten sie aus, um auf sie differenziert zu reagieren und auf die differenzierten Reizreaktionen wiederum differenziert reagieren zu können. Die differenzierten Reaktionen setzen ein reflektiertes Verhältnis zu den Reizen voraus, aber nicht immer oder nicht in ihren einfachen Formen ein autonomes Subjekt.

Warum sollen wir dann in der Befolgung konstruktiver Regeln überhaupt eine Form der *Rationalität* sehen, zwar nicht der allgemeinen, aber einer besonderen? Im Fall der Befolgung konstitutiver Regeln ist die Antwort klar: sie konstituiert ein wissenschafts- und handlungsfähiges Subjekt, das all die Fähigkeiten hat, die traditionell als die der Vernunft verstanden wurden. Sie schafft den Träger der Vernunft; daher kann man sogar in der Befolgung konstitutiver Regeln die Betätigung einer vor und unabhängig vom Subjekt bestehenden Vernunft sehen. Was aber haben Regeln, deren Befolgung die besonderen Wirklichkeitsbereiche der Wissenschaft, der Sprache, der Kunst und ihrer verschiedenartigen Gegenstände erschließen, mit Vernunft oder Rationalität zu tun?

Als vernünftig oder rational gilt alles Handeln und Urteilen, das Gründen folgt. Die besonderen Wirklichkeitsbereiche, die durch die konstruktiven Regeln erschlossen werden, liefern spezifische Bereiche von Gründen, die für Entscheidungen in anderen Bereichen irrelevant sind. Daher stecken sie nicht nur besondere Bereiche der Wirklichkeit, sondern auch der Rationalität ab. Gründe in der wissenschaftlichen Methodologie sind für die Begründung ästhetischer Entscheidungen irrelevant; Gründe für den Gebrauch bestimmter Wörter zur Beschreibung bestimmter Erfahrungen irrelevant für die Begründung von Entscheidungen in der Politik oder Ökonomie. Konstruktive Regeln zerfallen notwendig in Klassen, die verschiedene Sphären der Rationalität erschließen; Sphären mit eigener

Gesetzlichkeit oder Rationalität. Man könnte vielleicht darauf bestehen, daß die Begriffe der Vernunft und Rationalität hier nichts zu suchen haben; daß man nur von einer *Eigengesetzlichkeit* der Sphären reden dürfe. Aber das wäre ein unfruchtbarer terminologischer Streit. Entscheidend ist, daß konstruktive Regeln *Gründe* spezifischer Art liefern. Deren Befolgung eine besondere Rationalität zu nennen scheint mir völlig angemessen.

Kann also auch ein Wesen, das kein autonomes Subjekt in Kants Sinn ist, Gründen folgen? Das (nach Kant) subjektkonstituierende Urteilen stellt tatsächlich nur einen Sonderfall des Gründebefolgens dar. Nach Kant ist es das Urteilen, das das gewöhnliche Befolgen von Gründen erst ermöglicht. Aber dies Ermöglichen kann nur in einem sehr besonderen und jedenfalls nicht in einem kausalen und zeitlichen Sinn verstanden werden. Wenn wir die Entwicklung von Kindern betrachten, sehen wir, daß sie lernen Gründen zu folgen, wenn sie sprechen lernen, aber noch nicht verantwortlich sind, sobald sie sprechen können. Was muß hinzukommen? Die Verwandlung des reizgesteuerten Wesens in ein autonomes Subjekt scheint ein gradueller Prozeß, der erst dann das Maß erreicht, das erlaubt, dem Wesen Verantwortlichkeit zuzuschreiben, wenn ein größerer Bruchteil seiner Verhaltensweisen nicht mehr reizgesteuert, sondern gründegeleitet ist. Hier stellen sich viele Fragen, die ich offen lassen muß.

Dagegen ist ein Hinweis darauf angebracht, daß die Regeln, die ich *konstitutiv* nenne, keine Regeln sind, die John Searle in seiner *Construction of Social Reality* konstitutiv nennt. Searles konstitutive Regeln „create the very possibility of certain activities. Thus the rules of chess ... create the very possibility of playing chess. The rules are *constitutive* of chess in the sense that playing chess is constituted in part by acting in accord with the rules. If you don't follow at least a large subset of the rules, you are not playing chess."[15] Diese Erläuterung paßt genau auf die Regeln, die ich *konstruktiv* nenne. Denn die Befolgung konstruktiver Regeln ermöglicht immer *bestimmte* Tätigkeiten, „*certain* activities", wie beschreiben oder Kunstwerke schaffen. Was ich dagegen konstitutive Regeln nenne, dessen Befolgung ermöglicht keine bestimmten Handlungen, sondern zurechenbares oder rationales Handeln. Für diese Regeln fehlt Searle ein eigener Begriff.

Terminologische Fragen sind zwar großenteils willkürlich entscheidbar; mir scheint aber Searles Gebrauch von *konstitutiv* nicht glücklich. Denn wie er selbst sagt: Schachspielen wird nur „*in part* by acting in accord with the rules" konstituiert. Das Regelbefolgen genügt nicht; zwei Schauspieler etwa können Schachfiguren nach den Schachregeln bewegen, ohne Schach zu spielen. Der Inhalt von Searles konstitutiven (und meinen konstruktiven) Regeln ist daher nicht beliebig; er muß vielmehr dem angemessen sein, was ihre Befolgung zugänglich macht. Das ist im Fall methodologischer Regeln einer Wissenschaft offensichtlich: sie müssen so sein, daß die Wirklichkeit, die die Wissenschaft erforschen will, erkennbar wird. Die konstruktiven Regeln konstituieren hier zwar in dem weichen Sinn, in dem Searle von konstitutiv spricht, die Wissenschaft, aber nicht ihren Forschungsgegenstand, dem sie sich anpassen muß. Daher legen sie nahe, sie konstruktiv statt konstitutiv zu nennen. Im Fall von Schachregeln scheint man allerdings eher von konstitutiven Regeln sprechen zu müssen, weil das von ihnen ermöglichte Schachspiel nicht, wie die Wissenschaft ihrem Untersuchungsgegenstand, anscheinend noch einem anderen Bereich angemessen sein muß.

15. John R. Searle, The Construction of Social Reality, London: Allen Lane 1995, 27f.

Tatsächlich ist der Unterschied nur graduell. Denn auch das von den Schachregeln konstituierte Schachspiel muß bestimmten Erfordernissen angemessen sein, die unabhängig und vor dem Schachspiel bestehen. Es muß den Anforderungen der Menschen angemessen sein, die Vergnügen darin finden, ein Spiel von der Art des Schachs zu spielen. Es muß unterhaltsam sein wie Glücksspiele, sein Ausgang soll dennoch nicht vom Glück, sondern der Klugheit des Spielers abhängen; seine Regeln dürfen nicht so einfach sein, daß es schnell zur Gleichheit der Spielstärke kommt, wie bei Mühle oder Dame; aber auch nicht so kompliziert, daß man sie schlecht überschauen kann. Solche Erfordernisse lassen zweifellos Spielraum für eine größere Anzahl schachähnlicher Spiele, schließen aber Willkür in der Wahl der Regeln aus. Auch Schach- und sonstige Spielregeln sind daher konstruktiv in dem Sinn, daß sie helfen, eine unabhängig von ihnen bestehende, durch menschliche Neigungen oder sonstige Gegebenheiten bedingte und spezifische Forderungen stellende *Möglichkeit* zu verwirklichen.

Allerdings nehme ich nicht an, daß die von der Befolgung konstruktiver Regeln eröffnete Wirklichkeit notwendig sozial ist. Doch auch die Wirklichkeit, die Searles *constitutive rules* eröffnen, muß nicht sozial sein, obgleich sie die Konstruktion der *sozialen* Wirklichkeit erklären sollen. Denn wenn sie konstitutiv sind für das Schach, das zwei Spieler spielen, können sie auch für Patience konstitutiv sein, das man allein spielt.

Zu Humes und Kants Thesen

Für Hume ist, wie Jean Hampton feststellte, die Vernunft „a purely informational faculty, working out relations of ideas and causal connections"[16], die zwar in unsere Handlungsgründe eingehen können, aber keine reine praktische Vernunft ermöglichen. Aber selbst wenn seine Sicht der Vernunft zu eng war, muß er nicht auch die Moral zu eng gesehen haben. Mit der Unterscheidung von konstitutiven und konstruktiven Regeln läßt sich das Verhältnis, in das Hume Vernunft und Moral setzt, besser beschreiben als mit der Unterscheidung von instrumenteller und substantieller Vernunft. Wir können nun sagen: er spricht den Regeln der Moral die Eigenschaften der allgemeinen Rationalität ab, die bei Kant als die Rationalität konstitutiver Regeln auftritt und (noch bei Korsgaard) nicht als *allgemeine* von der *besonderen* Rationalität unterschieden wurde. Er spricht ihnen aber nicht die Eigenschaften konstruktiver Regeln ab, die die der besonderen Rationalität sind.

Nach Hume liefert nicht die Vernunft, sondern eine *passion*, die *sympathy*, die moralspezifischen Gründe, denen wir in moralischen Urteilen folgen. Diese Annahme schließt aus, daß die Moral ein Produkt der allgemeinen Rationalität und Normativität ist, aber nicht, daß sie ein Produkt der besonderen Rationalität und Normativität ist. Die *sympathy* liefert uns ebenso den Ansatz zu tugendhaftem Handeln und Urteilen wie eine natürliche Empfindungsäußerung den Ansatz zu verbalen Empfindungäußerungen. Wenn wir den konstruktiven Regeln folgen, die tugendhaftes Handeln und Fühlen und verbale Empfindungsäußerungen bestimmen, wird eine eigengesetzliche Sphäre der Wirklichkeit erschlossen, die der Moral und die der propositionalen Sprache.

16. Jean Hampton, On Instrumental Rationality, in J.B. Schneewind, Hg., Reason, Ethics, and Society, Chicago: Open Court, 1996, 84–116, 101.

Man gäbe von Humes Moraltheorie ein falsches Bild, wollte man aus seiner Aussage, die Vernunft solle nur die Sklavin der *passions* sein, ableiten, er halte moralisches Handeln für unvernünftig. Hume nimmt dessen *eigene* besondere Rationalität implizit an. Nur weil er sowenig wie Kant die allgemeine von der besonderen Rationalität unterschied, setzte er Rationalität mit der allgemeinen gleich. Vor die Frage gestellt, ob moralisches Handeln rational ist, zog er wegen der Eigenarten moralischen Handelns, in denen es nicht mit der allgemeinen Rationalität übereinstimmt, die Verneinung vor, während Kant wegen der Eigenarten, in denen es der besonderen Rationalität entspricht, die Bejahung vorzog. Hume konnte sich daher leicht als moralischen Irrationalisten mißverstehen, während Kant sich vor die harte Aufgabe gestellt sah, die Identität von Moralität und allgemeiner Rationalität nachzuweisen.

Mit dem Verständnis der Moral als eine besondere Wirklichkeitsdimension erschließend steht Hume nicht allein. Auch nach Aristoteles folgen wir, wenn wir tugendhaft sind, Gründen, die nicht nur die allgemeine Qualität haben, Handlungsgründe zu sein, sondern eine weitere Eigenschaft, nämlich zu sichern, daß unser Handeln unsere spezifischen Fähigkeiten betätigt. Auch für Aristoteles entwickelt moralisches Handeln eine Wirklichkeitsdimension, die nur durch eine zwar erklärbare, aber nicht begründbare und auch nicht begründungsbedürftige Entscheidung erschlossen wird; eine Entscheidung für die Betätigung der eigenen Fähigkeiten. In diesem Punkt steht Aristoteles Hume näher als Kant, trotz anderer Gemeinsamkeiten, die Korsgaard zwischen Aristoteles und Kant sieht.[17]

Nach Kant ist die Moral zwar nicht mit der Vernunft insgesamt identisch, sondern mit der *praktischen* Vernunft. Diese besteht im Befolgen von Gründen für *Handlungen*, nicht für Meinungen. Aber er schließt aus, daß die Gründe für moralische Entscheidungen noch durch andere Eigenschaften qualifiziert sind als die, Handlungs*gründe* zu sein. Es entspricht nicht seiner Erwartung an die Rationalität, die er dem moralischen Handeln und Urteilen unterstellt, daß in einer *passion* wie der *sympathy* oder einem materialen Ziel wie dem der Betätigung *bestimmter* Fähigkeiten Handlungsgründe angelegt sein könnten, die durch konstruktive Regeln entwickelt werden und spezifisch moralisch sind. Es könnte nun scheinen, daß Kant dabei nur der mangelnden Unterscheidung von allgemeiner und besonderer Rationalität zum Opfer fiel. Das ist jedoch nicht der Fall. Er hatte vielmehr einen guten Grund, der über den hinausgeht, daß moralisches Handeln nicht einfach irrational ist und daher der allgemeinen Rationalität entsprechen muß. Dieser Grund macht das Verhältnis von Vernunft und Moral besonders schwer zu durchschauen.

Der Grund ist, daß wir der Moral für die menschliche Existenz eine fundamentalere Rolle zuweisen, wenn wir sie mit der allgemeinen Normativität gleichsetzen, als wenn wir sie mit der besonderen Normativität gleichsetzen. Sie regelt das Handeln dann nicht nur in einem besonderen Wirklichkeitsbereich, dem der ordinären Moral, sondern in jedem möglichen. Eine solche fundamentalere Rolle müssen wir der Moral aber offenbar zuweisen. Denn nur so können wir der Tatsache Rechnung tragen, daß wir *jede* mögliche Entscheidung darauf befragen können, ob sie moralisch richtig ist.

17. In Korsgaard, From Duty and for the Sake of the Noble: Kant and Aristotle on Morally Good Action, in St. Engstrom und J. Whiting, Hg., Aristotle, Kant, and the Stoics. Rethinking Happiness and Duty, Cambridge UP 1996, 20–36.

Verstehen wir moralische Regeln mit Hume als Regeln, die einen besonderen Bereich erschließen, so wird es sinnlos, für Entscheidungen außerhalb dieses Bereichs zu fragen, ob sie moralisch richtig sind. Moralische Urteile betreffen dann nur den Bereich, der durch eine Definition der Moral von der Art, wie ich sie zu Beginn des Aufsatzes gegeben habe, festgelegt wird. Aber es *ist* durchaus sinnvoll zu fragen, ob wir irgendeinen beliebigen Bereich, etwa den, den die empirischen Wissenschaften erschließen, erschließen *sollen*. Mit einer solchen Frage gehen wir zwar über den Bereich der *ordinären* Moral hinaus, aber nicht über den Bereich sinnvoller moralischer Fragen. Wenn die Moral nur eine *bereichsspezifische* Normativität hat, kann man auch nur ein *bereichsspezifisches* Kriterium der normativen Richtigkeit angeben. Aber wir brauchen ein bereichs*un*spezifisches Kriterium der Normativität. Also scheint es, daß wir es nur in der bereichs*un*spezifischen, also allgemeinen Normativität und Rationalität finden können.

Für diesen Schluß scheint auch folgende Überlegung zu sprechen. Eine ordinäre Moral ist Teil der Kultur oder Lebensform einer Gesellschaft. Verschiedene Gesellschaften haben, wie die Geschichte lehrt, verschiedene Kulturen und Moralen gehabt. Wenn wir keine moralischen Relativisten sein wollen, müssen wir ein kulturunabhängiges Kriterium der moralischen Richtigkeit annehmen. Können wir es irgendwo anders suchen als darin, daß die richtige Moral die Moral einer Gesellschaft sein muß, die den Standards keiner besonderen, sondern der allgemeinen Rationalität entspricht? Vielleicht gibt es solche Standards nicht. Aber dann können wir offenbar alle Hoffnung fahren lassen, ein kulturunabhängiges Kriterium der moralischen Richtigkeit zu finden.

All diesen Überlegungen zum Trotz kann Kants Lösung nicht richtig sein. Denn seine These, die moralische Normativität sei die allgemeine Normativität und Rationalität, hat eine fatale Konsequenz. Sie setzt *verantwortliches* und *moralisch richtiges* Handeln gleich. Die Regeln, die uns als autonome Subjekte konstituieren, machen uns schon moralisch. Sobald wir nicht moralisch handeln, sind wir auf den Status reizreagierender Organismen zurückgefallen und können sowenig wie diese verantwortlich handeln. Wir sind nicht wir selbst, wenn wir böse handeln. Kurz: böses *Handeln* ist unmöglich. Diese kontraintuitive Konsequenz kehrt in Korsgaards Kantverteidigung wieder. Werfen wir daher einen Blick auf sie.

Korsgaards Kantverteidigung

Ich gliedere Korsgaards Kantverteidigung in sechs Schritte.[18]

1. Unser Vermögen der Reflektion verurteilt uns dazu, nach Gründen der Richtigkeit unserer spontanen und alternativen Handlungen und Meinungen zu suchen.
2. Die Reflektion verwandelt uns in ein Selbst, das ein Gesetz des Handelns braucht; sonst würden wir nur wieder auf Reize reagieren.
3. Um selbst zu handeln, dürfen wir das Gesetz keiner Neigung oder vorgefundenen Ma-

18. Vgl. v.a. Korsgaard, The Sources of Normativity, Cambridge UP 1996, und The Normativity of Instrumental Reason, in Garrett Cullity, Hg., Ethics and Practical Reason, Oxford 1997, 215–54. Das im Text folgende Argument ist diesen beiden Arbeiten entnommen. Vgl. bes. The Sources of Normativity pp. 220f.

terie anpassen. Wir müssen nach Maximen handeln, die wir zum Gesetz erheben können.

4. Dies Prinzip, Kants kategorischer Imperativ, ist zu formal, um als *moralisches* Gesetz zu gelten. Es kann auch einem konsequenten Verbrecher Identität geben.
5. Die Normativität der Moral ist dennoch dieselbe Normativität, die sich in der rationalen Notwendigkeit des Logischen darstellt.
6. Wir folgen dem Sittengesetz, wenn wir nur solchen Maximen folgen, die dem Handeln *aller* Vernunftwesen das Gesetz geben.

Betrachten wir die Schritte im einzelnen. Der *erste* Schritt, die Annahme, daß unser Reflektionsvermögen uns nach Gründen für unsere Handlungen und Meinungen suchen läßt und wir in der Befolgung von Gründen der Normativität der Vernunft folgen, scheint mir unanfechtbar. Vielleicht wäre es besser, das Reflektionsvermögen Vernunft zu nennen, weil das bloße Bewußtsein einer spontanen Handlung noch nicht zur Suche nach Gründen führt. Aber dieser Punkt rührt nicht die Richtgkeit des ersten Schritts an.

Daß uns die Reflektion in ein Selbst verwandelt, wie der *zweite* Schritt annimmt, ist ebenso unanfechtbar. Aber braucht ein Selbst ein *Gesetz* des Handelns, um nicht nur auf Reize zu reagieren? Korsgaards Antwort lautet: „I cannot regard myself as an active self, as *willing* an end, unless *what I will* is to follow my maxim in spite of temptation ... if I am to regard *this* act, the one I do now, as the act of my *will*, I must at least make a claim to universality, a claim that the reason for which I act now will be valid on other occasions, or on occasions of this type"[19]. Wenn man nach Gründen entscheidet, entscheidet man unter der Bedingung, daß die Gründe nicht nur die vorliegende Entscheidung rechtfertigen, sondern jede vergleichbare. Man kann nur nach Gründen entscheiden, wenn man universalisierbaren Regeln folgt.

Korsgaards Antwort ist richtig. Man kann an der Terminologie mäkeln und zweifeln, ob man sagen sollte, daß ein Selbst ein *Gesetz* des Handelns braucht. Man kann auch beanstanden, daß nicht genügend geklärt ist, was Universalisierbarkeit heißt. Solche Zweifel sind aber unerheblich für den entscheidenden Punkt: daß ein Selbst Handlungsfähigkeit verliert, wenn es nicht *irgendwie* universalisierbaren Regeln folgt.

Dies Argument reicht auch aus, um den *dritten* und *vierten* Schritt zu rechtfertigen. Nach dem dritten Schritt handeln wir nach dem kategorischen Imperativ, wenn wir Gründen folgen; nach dem vierten Schritt ist der kategorische Imperativ nicht das Sittengesetz, weil er nur eine Konsequenz des Handelns verlangt, zu der auch ein Verbrecher fähig ist.

Soweit hat Korsgaard viel gezeigt: daß der kategorische Imperativ ein *konstitutives* Prinzip ist, und daß er ein handlungsfähiges und zurechenbares Subjekt konstituiert. Aber sie beweist nicht, was hier in Frage steht: daß reine praktische Vernunft oder der kategorische Imperativ *moralisch*, ja das *Sittengesetz* ist. Sie nimmt vielmehr an, daß der kategorische Imperativ keine spezifisch *moralischen* Handlungen auszeichnet, sondern nur vernünftige.

Wie aber kommt man allein auf dem schmalen Pfad der rationalen Notwendigkeit vom kategorischen Imperativ in Korsgaards magerem Sinn zum Sittengesetz in Kants fettem Sinn?

Korsgaard beruft sich dafür auf Wittgensteins Kritik der Möglichkeit einer Privatsprache[20]. Wenn ich es richtig verstehe, ist ihr Argument dies: Private Gründe kann es sowenig geben wie private Zeichenbedeutungen. Das Gesetz, das sich ein Verbrecher gibt, ist ein privater Handlungsgrund, öffentliche Handlungsgründe dagegen sind notwendig moralisch.

Dies Argument scheint mir unakzeptabel. Erstens sollte eine Moraltheorie nicht auf anfechtbare Theorien wie die Wittgensteinsche über die Möglichkeit privater Zeichenbedeutungen zurückgreifen müssen. Zweitens und wichtiger läßt das Argument im Dunkel, warum das Gesetz des Verbrechers ein *privater* Handlungsgrund oder warum öffentliche Handlungsgründe notwendig moralisch sind. Korsgaard scheint hier selbst Zweifel zu haben. Sie betrachtet den Fall des konsequenten Verbrechers, der dem Gesetz der Verbrecherehre folgt. Sie sagt: Wenn er

> „attempted to answer the question why it matters that he should be strong and in his sense honour-bound even when he was tempted not to, he would find that its mattering depends on the value of his humanity, and if my other arguments go through, he would find that that commits him to the value of humanity in general, and so to giving up his role as a Mafioso. But suppose – as is likely enough! – that he never does work all this out? Where does that leave him? ... there is a real sense in which you are bound by a law you make for yourself until you make another ... There is a sense in which these obligations" – von Mafiosi und andern Kriminellen – „are real – not just psychologically but normatively ... I know that this conclusion will seem outrageous to some readers. ... The point is just this: if one holds the view, as I do, that obligations exist in the first-person perspective, then in one sense the obligatory is like the visible: it depends on how much of the light of reflection is on."[21]

Das heißt: Man kann das Gesetz, das man sich gemäß dem mageren kategorischen Imperativ geben muß, um autonom zu sein, nur vom je eigenen Standpunkt wählen. Aber auch das *Sittengesetz* kann man immer nur vom eigenen Standpunkt wählen, wenn man dabei auch erstens den Wert der eignen Menschheit und zweitens den der Menschheit überhaupt anerkennen sollte. Zugleich muß es immer auch das Gesetz sein, das Autonomie und Verantwortlichkeit erst ermöglicht. Deshalb ist es nach ihr wie nach Kant unmöglich, daß man zugleich verantwortlich und unmoralisch handeln kann. Wenn sie dem Mafioso nicht die Verantwortlichkeit für sein Handeln absprechen will, was sie nicht tut, bleibt ihr nur die Konsequenz, ihn auch dann als moralisch anzuerkennen, wenn er nach dem Gesetz des Mafioso handelt. Denn dies konstituiert seine Autonomie.[22]

19. Korsgaard, The Sources of Normativity a.a.O. 231f.

20. Ebd. 136ff.

21. Ebd. 256f. Korsgaards Argument schließt auch den Verweis auf eine als privat verstandene Empfindung als einen Handlungsgrund aus, etwa darauf, daß mein Schmerz mich rechtfertigt, ein heißes Eisen fallen zu lassen. Korsgaard verteidigt sich ebd. 147 so: „I am not denying that when we are in pain part of what is going on is that we are having sensations of a certain character. I am however denying that the painfulness of pain consists entirely in the character of those sensations ... Pain wouldn't hurt if you could just relax and enjoy it." Sie behauptet auch, ebd. 150: „Pain is the *unreflective* rejection of a threat to your identity. So pain is the *perception* of a reason, and that is why it seems normative."

22. Ebd. 257 legt sie sich nicht darauf fest, die Verpflichtung, der Verbrecherehre treu zu bleiben,

Diese Konsequenz ist jedoch nicht nur *outrageous*. Sie ist eine *demonstratio ad absurdum* der Prämissen. Wenn diese den Schluß erlauben, jemand könne wegen der Besonderheit des Standpunkts, auf dem er seinem Handeln ein Gesetz gibt, zu *Verbrechen verpflichtet* sein, sollte man mindestens eine von ihnen verwerfen statt den Schluß anzuerkennen. Es liegt nahe, da wir die ersten vier Schritte anerkannt haben, den *fünften* Schritt in Korsgaards Kantverteidigung zu verwerfen, die Annahme, die Normativität der Moral sei dieselbe Normativität wie die des Logischen.

Deren Verwerfung liegt auch deshalb nahe, weil Korsgaard selbst als Bedingung des Schritts zum Sittengesetz hervorhebt, daß wir „the value of humanity in general" anerkennen, sogar daß wir „value anything at all"[23]. Diese Anerkennung ist eine Wertentscheidung, die über die *Konstitution* des handlungs- und zurechnungsfähigen Subjekts hinausgeht. Sie ermöglicht die *Entfaltung* einer eigenen Wirklichkeitsdimension durch *konstruktive* Regeln, als die auch Korsgaard die Regeln der Moral hier implizit anerkennt.

Ein Rückgriff auf Leibniz' Moraltheorie

Wenn wir daran festhalten, daß wir auch für moralisch falsches Handeln verantwortlich sein können, kann die Normativität der Moral nicht mit der allgemeinen Normativität identisch sein. Wie können wir dann aber der Tatsache Rechnung tragen, daß wir für Entscheidungen nicht nur in einem besonderen Wirklichkeitsbereich, sondern für jede beliebige Entscheidung fragen können, ob sie moralisch richtig ist?

Es reicht offenbar nicht aus, moralische Regeln als konstruktiv für einen besonderen Bereich zu beschreiben. Sie müßten auch Kriterien dafür liefern können, unter welchen Bedingungen ein neuer Bereich überhaupt erschlossen werden sollte. Dieser Bedingung entspricht Leibniz' Moralverständnis. Leibniz' höchstes Moralprinzip ist die Regel, jede Möglichkeit zu verwirklichen, soweit sie nicht die Maximierung verwirklichbarer Möglichkeiten vermindert. Daher erklärt er eine Welt mit den meisten kompossiblen Sachverhalten für die beste aller möglichen[24] Sein Moralprinzip liefert auf die Frage, wann wir eine Wirklichkeitdimension erschließen sollen, die Antwort: immer, solange die Gesamtheit der Wirklichkeit dadurch nicht vermindert wird. Das ist eine konstruktive Regel, weil sie von vorausgesetzten verantwortlichen Subjekten verlangt, gegebene Möglichkeiten in einer bestimmten Weise zu verwirklichen.

Ich kann hier nicht untersuchen, wie weit eine solche Moralprinzip trägt. Ich gebe es nur als ein Beispiel dafür an, daß man moralische Regeln als konstruktiv verstehen und doch die Frage beantworten kann, auf die Hume keine Antwort hat.[25] Es würde auch ein *kulturunabhängiges* Kriterium der moralischen Richtigkeit liefern. Eine Gesellschaft ist nach ihm in dem Maß rational oder moralisch, wie sie kompossible Möglichkeiten verwirklicht. Allerdings könnte ein Kantianer sich durch den Leibnizschen Ansatz bestätigt sehen. Daß wir alle kompossiblen Sachverhalte verwirklichen sollen, könnte er als ein Gebot der Vernunft in ihrer Rolle begründen, die möglichste Vollständigkeit ihrer Entscheidungsgründe und daher eine möglichst vollständige Erforschung der Wirklichkeit zu fordern. Diese Forderung würde man verletzen, wenn man nicht alle kompossiblen Sachverhalte verwirklichte. In der Verfolgung des Leibnizschen Prinzips folgen wir, so könnte der Kantianer sagen, der Vernunft allein, der *reinen* Vernunft.

Könnte sich der Kantianer nicht noch weiter in seiner Gleichsetzung von Rationalität und Moralität bestätigt sehen? Denn mag auch das Leibnizsche Moralprinzip eine konstruktive und keine konstitutive Regel sein, der Kantianer kann so argumentieren: ‚Das Leibnizsche Kriterium ist nicht das der überlieferten Moral. In dieser folgen wir nur spezifischen Kriterien einer tradierten Moral. Das allgemeine Kriterium kann daher nur das der allgemeinen, nicht der besonderen Rationaltiät sein; wenn wir überhaupt daran festhalten wollen, daß das Kriterium normativ richtigen Handelns den Bedingungen der Rationalität genügen soll.'

An dieser Bedingung sollte man allerdings festhalten. Dennoch ist das Argument nicht schlüssig. Es übersieht eine Konsequenz der Unterscheidung von allgemeiner und besonderer Rationalität. Sie macht eine weitere Unterscheidung notwendig. Wenn wir einmal anerkennen, daß Rationalität nicht nur in der Befolgung solcher („konstitutiver") Regeln besteht, die ein zu Wissenschaft und Verantwortung fähiges Subjekt konstituieren, sondern auch solcher („konstruktiver") Regeln, die besondere Wirklichkeitsdimensionen erschließen, dann können wir auch nach Eigenschaften und Bedingungen der Befolgung konstruktiver Regel und nach Gemeinsamkeiten der Formen der besonderen Rationalität fragen. Genau das tun wir, wenn wir nach einem allgemeinen Kriterium der normativen Richtigkeit *jedes* Handelns fragen, nicht nur des Handelns in solchen Bereichen, für die die tradierten Moralen ihre Richtigkeitskriterien angeben. Wir suchen dann nach einer Bedingung, der wir alle möglichen Formen der besonderen Rationalität unterwerfen sollten. Wir können eine solche Bedingung eine allgemeine Bedingung der Formen der besonderen Rationalität oder kurz eine allgemeine Rationalität nennen. Aber sie wäre nicht die allgemeine Rationalität in Kants Sinn, in dem sie eine Bedingung der Konstitution des Subjekts ist.

Wenn wir also Moralität und Rationalität am Ende doch identifizieren, wie man es durch die Annahme des Leibnizschen Kriteriums impliziert finden kann, dann identifizieren wir Moralität nicht mit der allgemeinen oder subjektkonstituierenden Rationalität, die Kant – und Korsgaard und die philosophische Tradition vor Kant – als Rationalität verstehen, sondern mit der Rationalität, der alle Formen der besonderen Rationalität genügen müssen. Sie hätte vermutlich auch Hume als Rationalität anerkennen können, obgleich er sie sowenig vom traditionellen Vernunftbegriff unterschied wie Kant.

Die Antworten

Zurück zu den Titelfragen. Ich will sie so beantworten. Die Vernunft ist praktisch, wann immer wir Handlungs*gründen* folgen. Der *reinen* praktischen Vernunft folgen wir, wenn wir einem Handlungsgrund folgen, den nur die Vernunft setzt. Ob wir das je tun, lasse ich

moralisch zu nennen, wohl aber darauf, daß sie eine Verpflichtung ist, die der Verbrecher erfüllen sollte.

23. Ebd. 256 und 125.
24. Leibniz, Philosophische Schriften, ed. Gerhardt. Bd. 5, Berlin 1882, 286.
25. Vgl. U. Steinvorth, Warum überhaupt etwas ist, Reinbek: Rowohlt 1994, 105-47.

offen. *Moralisch* im Sinn der ordinären Moral ist die Normativität dann, wenn sie eine besondere Dimension der Wirklichkeit entwickelt; im Sinn einer reflektierten Moral dann, wenn sie ein Kriterium dafür liefert, wann eine Dimension entwickelt werden soll. In beiden Fällen sind ihre Regeln konstruktiv. Im zweiten Fall könnten Handlungsziele begründet werden, die man als solche der reinen Vernunft verstehen kann.

Wichtiger als diese Antworten ist die Unterscheidung in drei Arten der Rationalität und Normativität, die zur Beantwortung der Fragen nötig war: in die allgemeine, subjektkonstituierende Rationalität, die besondere, einen Wirklichkeitsbereich eröffnende Rationalität und die Rationalität, die eine Bedingung der Formen der besonderen Rationalität ist.

Schemata, Abstraction, and Biology
Man as the Abstract Animal rather than the Symbolic Species?

FREDERIK STJERNFELT

"... the brutes use signs. But they perhaps rarely think of them as signs."
(Peirce CP 5.534 (1905))

Against the background of the role of schemata in biosemiotics and of the current theories of complex adaptive systems, Terrence Deacon's claim that man can be characterized as the "symbolic species" is critically revised. With reference to Deacon's use of Peirce's semiotics, it is pointed out that Peircean symbols occur at much more basic levels in biology than at the missing link threshold. Instead, Peirce's diagram and abstraction theories are invoked in a proposal for a better description of the animal-man distinction.

With the mapping of the human genome coming close to completion, there is an increasing interest in higher levels of organization in biology. Complexity theory and biosemiotics are two different traditions which, each in their own way, highlight the relations between biology, meaning, and language. The Santa Fe tradition in complexity theory has, during the last decade, addressed the issue of the formal complexity of natural and artificial systems of many substantially different kinds. Its main idea is that such systems have two crucial characteristics:

1. They are complex, that is, their large-scale behavior is the result of the interaction of a multitude of connected units on a lower scale, and this behavior is non-linear with respect to environmental stimuli.
2. They are adaptive, that is, they possess the possibility of stably modifying themselves as a result of environmental pressures – hence the nickname CAS, "complex adaptive systems", for the object of the Santa Fe institute's research, covering a wide range of systems including physics, economics, sociology, and, prototypically, biology.

Now, many CASes develop internal representations of environmental information and this has led the Santa Fe tradition to take up the age-old philosophical issue of *schemata*, sketch-like iconic representations uniting economy, generality, and malleability.[1] Such schematic representations permit some degree of prediction based on environmental information and thus give rise to successful action, that is, adaptability. Such schematic competence seems to be a crucial property in at least a wide-range of sufficiently complex CASes, including all higher animals (that is, animals equipped with a central nervous system).

Biosemiotics is another recent tradition taking up similar facts from a different point

1. Cf. for instance the inclusion of Ben Martin's introductory paper "The Schema" in the Santa Fe Institute's *Proceedings* vol. XIX (Cowan et al., 1994).

B. Brogaard, B. Smith (Hg.), *Rationality and Irrationality / Rationalität und Irrationalität*, S. 341–361.

of view. Biosemiotics takes at face value the existence of spontaneous semiotic vocabulary in biology (genetic information, genetic code, RNA-messenger, etc.) and investigates biological processes at all levels from a semiotic point of view.[2] This implies, of course, that the enterprise may develop the core concepts of semiotics in new directions (a widespread semiotic assumption like the sign's dependency on consciousness may, for instance, need to be relativized). The idea is that the spontaneous semiotic vocabulary present in biology does not owe its existence to accident but reflects a core part of biological ontology. Hence a purified version of semiotics may constitute a crucial part of biological ontology, and a main task of biosemiotics will be to cultivate a more sophisticated use of semiotic terminology in order to make clear the a priori systems of ontological concepts used in biology.[3] This entails, moreover, the possibility of distinguishing between different semiotic objects and processes in biology – presumably resulting in a spectrum going from the simplest sign types in primitive life forms to ever more complex types during the course of evolution.[4] My idea is that the simplest semiotic process is probably categorical perception as the behavioral discretization of a continuum, a process of a type to be found already in bacteria.[5]

Both these conceptual revolutions are still in the making; but it is evident that they imply the possibility of asking the old question of the transition from ape to man in a quite new way. If biological evolution is characterised by – among other properties – increasing semiotic sophistication, then it becomes an urgent question to ask which decisive semiotic change took place, if any, with the introduction of language during the Homo Habilis period in human prehistory: what is the semiotic missing link?

In this context, Terrence Deacon's much-discussed and groundbreaking book *The Symbolic Species* attempts a new answer. Based on a threefold background of arguments – philosophical, neurological, and anthropological, – Deacon proposes that the main event in the animal-man transition is the introduction of *symbols*. Now, as is well known, the concept of symbol is probably one of the most ambiguous notions in the history of thought, and Deacon takes great care to make precise the version of it he finds central. He picks the notion of symbol found in Charles Peirce's semiotics which stands apart from many other symbol concepts by assuming the symbol as a complex derivative notion related to simpler sign types included in its composition. In Peirce, the symbol presupposes the existence of the simpler sign types icon and index, respectively, so that it makes sense to say that genuine symbols are built up from icons and indices. Icons are signs defined by similarity to the object they refer to; indices are signs defined by actual connection to the object. Symbols, finally, are defined as signs referring to their – general – object by means of habit. These three classes are not mutually exclusive partitions of the field of signs; rather higher sign types presuppose and include more simple types. Thus, icons form the most fundamental sign type (with respect to object reference), and all higher types presuppose icons. Thus, indices are only possible in so far as they possess iconic

2. This has been pointed out by Emmeche and Hoffmeyer 1992.
3. As I have argued in Stjernfelt 1999.
4. This idea is proposed in Hoffmeyer 1996.
5. I put forward this idea in Stjernfelt 1992; I only recently discovered that exactly the same idea is to be found in Giorgio Prodi's writings from the 80's (e.g. Prodi 1988).

qualities: the footprint in the sand is a prototypical index, in so far as the sign refers to the object having caused it, but it also possesses iconic qualities, in so far as the footprint's shape to some extent resembles the shape of the foot which made it.

In order to interpret something as indexical, so Deacon, a higher-order relation must hold between two groups of icons (in the footprint case, there must be a relation between the group of possible footprints similar to the actual print on the one hand and the group of possible foot shapes responsible for it on the other). This corresponds, Deacon argues, to conditioned response in ethology. Furthermore, in order to construct a symbol, a whole group of indices are related by means of indexical relations between their tokens.[6] This relation, internal to the symbol, is now strengthened at the expense of the object reference of the initial indices. Thus, the symbol is as a tendency loosened from the closer object contact found in icons and indices. They are bracketed, allowing the symbol to function on its own in representation and reasoning, but in any specific interpretation of a symbol, its iconic and indexical basis must be reinserted, including the possibility of new iconoindexical specifications of it.

This forms the base of a fertile criticism of rival accounts which conceive symbols as atomic primitives, as for instance the physical symbol systems hypothesis, which makes symbols simple physical units corresponding in a rather direct way to other physical units. With reference to a long range of neurological brain scanning experiments which we shall not go into here, Deacon argues that philosophical points of view related to simpler symbol conceptions lack empirical support. This criticism is aimed especially at Chomsky's transformation grammar and related positions, which claim the existence of an innate grammar module in human beings, a module which allegedly should be completely lacking in apes. Deacon's scanning experiments point to the fact that sufficiently complex linguistic tasks inevitably give rise to very widespread brain activity including several separated parts of the cortex – which is a strong argument against the language module hypothesis. To counter it, Deacon claims a symbol hypothesis. Symbol use is taken as the distinctive advantage of mankind in comparison to other higher animals. Symbol use – as a complex phenomenon – naturally involves the integration of a large amount of highly differentiated, more primitive brain competences. Hence, such a hypothesis makes the semiotic animal-man transition more continuous so that symbol use would merely be the integration of a series of competences already to a large extent present in higher animals. The neural equivalent to symbol processing is neither a specific module of the brain nor the simple size difference of the cortex – but the *degree of integration* of the human brain, the latter being even neurologically measurable in terms of a much longer growth period and a larger degree of neuron interconnection between spatially distant parts (so as for instance the cortex and the cerebellum, very important for the automatization of phonetic aspects of speech).[7]

6. It should be noted that this recursive definition of symbols does not correspond to Peirce's original account. In Deacon, indices are made out of icons plus icon holding between icons, and symbols, in turn, out of indices plus indices holding between indices. In Peirce, however, the three are irreducible, and the icon-index structure rather forms the internal anatomy of the symbol without being sufficient for its compositional definition.
7. Deacon's strong neurological and semiotic arguments for his symbol missing link hypothesis, is, as far as I can see, not as devastating for Chomskyism as he presumes. Deacon's argument is

Deacon proposes both a general and a specific scenario for the evolution of speech. The specific one features complex anthropological hypotheses[8] which may be set to one side in this context. The more general scenario revives the notion of "Baldwinian" evolution (after the American psychologist James Mark Baldwin from around 1900). Baldwin proposed the idea that seemingly Lamarckian hereditary properties could be explained within a Darwinian framework with a support hypothesis: that animals able to learn new behaviors may be able to direct evolution as a result of that behavior – because the behavior in question will force fellow species members to assume it or perish. In such cases, a huge selection pressure will favor individuals most able to learn that behavior: "... *those congenital or phylogenetic variations are kept in existence which lend themselves to intelligent, imitative, adaptive, or mechanical modification during the lifetime of the creatures which have them.*" (Baldwin 1902, p. 95). This Baldwinian evolution argument is evidently stronger, the more intelligent the organism in question is. Thus, Deacon's idea is that the passage to symbol use is intimately connected to the fast evolution of the human brain during the last few millions of years. He imagines a scenario, in which embryonic symbol use in small humanoid groups kick-starts a process selecting for higher brains with symbol processing capabilities within those groups, thus speeding up evolution's pace dramatically, with our sophisticated symbol abilities as a result.

My aim here is to try and render the semiotic aspects of Deacon's hypothesis – with which I basically sympathize – somewhat more precise. The main problem is that even the Peircean definition of "symbol" is probably much too primitive – as well as too general – to explain the semiotic aspects of the animal-man transition. In addition, there are some terminological problems to be sorted out. For while Deacon claims that he uses Peirce's terminology, as a matter of fact he undertakes his own reconstruction thereof, changing the higher sign types – indices and symbols – in two ways. They are rendered compositional with respect to lower sign types, perhaps for reasons of theoretical economy; and they are rendered more complex than is the case in Peirce's account, perhaps in order to make them "fit" the ape/man boundary better. Thus, Deacon's explanation of the index seems to cover cases which, on Peirce's account, would automatically be counted as symbols. Pure indices – in Peirce only possible as a limit case – will be tied to the actual here and now, while it is the privilege of the symbol to possess an *esse in futuro* and thus form a habit, regulating future behavior. But when Deacon claims that conditioned behavior is

that a universal grammar competence is too abstract to be selected for, but it is not evident that this universal grammar could not be the result of logical inference principles with a high survival value. Moreover, elsewhere Deacon does not hesitate to ascribe survival value to rather general competences.

8. So as for instance that the increase in brain size necessitating more protein made early man turn to a more carnivorous behaviour. While the *Männerbund* went hunting, the mothers nursing the children were waiting for protein to be brought home. The hunting man's gene pool was threatened, however, by his woman's possible unfaithfulness during hunt, and she and her child was correlatively threathened by protein undernourishment if he did not return. This situation calls for stabilization by marriage which in turn requires stable institutions guaranteed by language - in turn calling for (further) development of symbol use. This hypothesis is interesting indeed, but it includes many specific issues and premisses which is not our concern in this context.

indexical, it seems he already includes future regulating features in indices (which are nowhere apparent in the prototypical footprint index). The problem is, of course, that if we accept the ordinary Peircean notion of symbol, then symbolic behavior becomes widespread in higher animals,[9] and then the notion becomes unfit for the task of distinguishing between animal and human behavior. Thus, a case of simple Pavlovian conditioning whereby the ringing of a bell releases the excretion of saliva in dogs involves a full-fledged symbol in Peirce's terminology: it is a habit, a regulation of future behavior, and it connects a continuum of possible bell sounds with a continuum of possible eating situations. Accordingly, Peirce's symbol concept includes a wide range of subtypes of very different complexity degrees, ranging from simple terms through propositions to whole arguments – each of these, in turn, including a whole fauna of further subtypes. Thus symbol use is neither as simple as Deacon presupposes (with respect to symbol subtypes) nor as complex as he presupposes (with respect to the simpler sign types)[10].

Deacon consequently adds some further requirements to Peirce's symbol concept in order to make it approximately fit the animal-man transition (to be sure, he maintains that a few higher animals, mostly apes, may learn simple symbol systems[11]). What he adds is a Saussure-like systematicity by requiring the co-presence of several interlinked symbols, both paradigmatically (implying the systematic difference between selected expressions) and syntagmatically (implying the syntactical organization of combined expressions). Only a system of this kind, he argues, permits ape and man to skip icon- and index-consciousness and their tight connection to the actual world and indulge in the semi-autonomous world of symbolicity in such a way as to give rise to the possibility of systematical counterfactual imagination. But as to systematicity, higher animals do possess taxonomies, both in perception (prey types) and communication (warning calls for different predators) so it seems systematicity is not the only key to the problem.

Thus it is correct that Peirce's symbol definition is a necessary, yet not sufficient, prerequisite for the construction of counterfactual possible worlds. The symbol notion seems too weak to account for the specific advantage of human semiotics over animal semiotics. Some more specific distinction within the field of symbols must be responsible for this decisive jump. Peirce himself only rarely considers the question of animal semiotic behavior, but he is, at least, quite sure that animals' abilities are far more elaborate than for instance those involved in simple conditioning. Take this late passage from around 1911:

9. It may even cover lower animals as well, cf. E. coli's ability to swim upstream in a saccharine gradient which in Peirce's terms must be classified as symbolic with respect to its *esse in futuro*. See Harnad 1987 for a long range of investigations of categorical perception tied to behaviour - hence forming symbols – in many different species.

10. A further problem is that Deacon's reconstruction of the icon-index-symbol triad as referred above makes it compositional, so that higher sign types are presumed reducible to combinations of the lower. But if pure icons – so Peirce – are mere possibilities, taken by themselves, then the actual dimension of indices can not be composed of ever so many icons; correlatively, the future dimension of symbols can not be composed of ever so many actual moments. Peirce's description face the opposite direction: symbols are wholes, and icons and indices are moments of the symbol's anatomy.

> Some seventy years ago, my beloved and accomplished school-ma'am taught me that human kind, being formed in the image of our Maker, were endowed with the power of Reasoning, while "the animals", lacking that power (which might have made them dissatisfied), received, each kind, certain "instincts" to do what was generally necessary for their lives. At least, so I understood her. But when I subsequently came to observe the behaviors of several big dogs and little birds and two parrots, I gradually came to think quite otherwise. For, in the first place, I gradually amassed a body of experiences which convinced me that many animals, perhaps all the higher ones, do reason, if by Reasoning is meant any mental operation which from the putting together of two believed facts leads to a Belief different in substance from either of those two. Once, for example, while I was driving (...) along a country road that was very familiar to me, a setter-dog that I had never seen before raced past me at the top of his speed. In an instant a turn of the road hid him from my sight. "Poor fellow!" I thought, "he races after his master in fear of losing him forever." A moment later, reaching the turn myself, I saw the dog again, not far ahead of me, but at a point where the road branched, and now sitting on his haunches. He was not panting nor showing the least sign of fatigue, but evidently puzzled which branch of the road. After a second or two, he started off at the same tremendous pace as before, on the more travelled of the two roads, though being the older and harder, it was not very obviously the more travelled of the two. These alternations, – a halt between two utmost speedings, with no slightest symptom of fatigue, – seemed to me to show plainly that the dog had stopped to consider which of the two branches of the road his master had probably taken; and his sudden choice of the more travelled showed that he concluded that his master would probably do as most people, which was a kind of argument: technically called a "probable deduction", – the commonest reasoning of a general in a campaign, when information is lacking, defective, or conflicting.
> (Manuscript 672, p. 2–5)

Peirce continues with another example of a parrot fooling a dog named Spitz. Every day, the dog's master would come home and call "Spitz, Spitz, Spitz!" in order to take the dog for a walk. If, by chance, somebody else came to the door, the parrot would repeat exactly the same yell, now provoking the dog to run to the door – only to be laughed at by the parrot who was presumably making a practical joke. It is of course very difficult to ascertain the amount or character of the reasoning taking place in animals from observation of behavior, but especially the first of Peirce's examples is illustrative of a type of deliberate choice that is seemingly widespread in higher animals. If we take this observation as being correct, this implies that higher animals do not only have at their command symbols in general, but also those most demanding and complex of symbols called arguments and reasoning involving diagrams. Even if not making it explicit, the dog's reasoning must implement a Y-shaped diagram in some fashion or other, making it possible for the dog to

11. Deacon refers at length to the famous *Kanzi* case where a young ape on its mother's back learned the symbolic language which scientists were trying to teach its mother. There is little doubt that *Kanzi* is a symbol user, both in Deacon's and (less surprising) in Peirce's sense of the word.

reason about which branch of the Y to choose. We may note, moreover, that the dog's situation at the fork in the road is also, at least in germ-like form, such as to contain the construction of another possible counterfactual world (like "what if my master had gone the other way ..."). If the more complex parrot example is to be accepted at face value, it must even be seen to contain the deliberate construction of a possible world *for* another animal, implying a theory of other minds – a type of behavior which seems well-documented in apes wanting to fool fellow apes and thereby draw them away from food, sexual partners, and so on, but let us stick to the more easy-to-interpret first example. Of course, the systematic exploration of worlds of alternative possibilities requires a stable representation system probably in the form of interconnected symbols – but the example here goes at least as far as to show that there is probably no upper bound to the complexity of symbol types which higher animals have access to as single signs. Animals do reason, and one could probably find cases displaying the use of both abductive, deductive, and inductive arguments, to take Peirce's own typology. A full-blown process of reasoning involves all three of them, and if animals reason, they master all of them – even if not explicitly, of course. They may guess, infer, and generalize from experience, respectively; they are rational because, just like us, they are forced to be – a Popperian argument.

So, the problem must lie elsewhere. The problem simply does not seem to lie in the degree of complexity of single symbol types, at least when measured on the term-proposition-argument complexity scale. Deacon is probably on the right track when he looks for the coming into place of a systematic interconnection of symbols making it possible to construct, evolve, and research in stable fashion possible worlds differing from the world that is actually perceived. But what makes the jump from a sophisticated reasoning ability and to a system of symbols possible? Deacon does not go into this question, but I think the explanation might be found in the ability to make signs explicit and undertake explicitly controlled reasoning with them – this pointing to another part of Peirce's work, namely his *abstraction theories,* which focus upon the possibility of making explicit the meaning of a term in stable fashion.

Peirce developed no less than two abstraction theories since he found the colloquial use of the word "abstraction" at his time (and in ours probably as well) to refer to two separate and autonomous problems. Both are relevant here. One is the mind's focusing ability, the other its ability to make problems explicit – referred to as distinction and hypostatic abstraction, respectively[12]. As often in Peirce, the two notions stem from different traditions in the medieval scholastic semantics.

Even if the abstraction problem recurs over and over in Peirce's work, he never consecrates a whole paper to unfold it, so an exposition must be based on a series of small notes spread in his published and unpublished work. The abstraction problem surfaces as early as 1867, but Peirce's interest in it reaches a peak in the fertile years of his mature theory of signs developed in the first ten years of the twentieth century. The first part of the the-

12. Both must, furthermore, be distinguished from *induction* dealing with a series of related, empirical phenomena and proposing a probable law uniting them. Induction is often by empiricists confused with abstraction, but like Husserl (2nd Logical Investigation), Peirce keeps these problems apart, and neither of the abstraction types have anything to do with extracting regularities from a set of examples.

ory, however, is stated as early as in "On a New List of Categories" (1.549; 1867)[13] where the trichotomy of *dissociation, prescission,* and *discrimination* is terminologically fixed. The idea is that there are three modes of separation which may be undertaken in the analysis of a phenomenon, going from the most coarse – being able to distinguish different qualities, e.g. red from blue (dissociation), through being able to distinguish what may be supposed to exist without the other, e.g. space from color (prescission), to the most subtle – being able to distinguish what may only be thought of separately, e.g. color from space (discrimination). This terminology remains constant in Peirce, and in "Syllabus" (1903), the three modes are directly connected to the definition of his three categories:

> In order to understand logic, it is necessary to get as clear notions as possible of these three categories and to gain the ability to recognize them in the different conceptions with which logic deals. Although all three of them are ubiquitous, yet certain kinds of separations may be effected upon them. They correspond to the three categories. Separation of Firstness, or Primal Separation, called *Dissociation*, consists in imagining one of the two separands without the other. It may be complete or incomplete. Separation of Secondness, or Secundal Separation, called *Prescission*, consists in supposing a state of things in which one element is present without the other, the one being logically possible without the other. Thus, we cannot imagine a sensuous quality without some degree of vividness. But we usually *suppose* that redness, as it is in red things, has no vividness; and it would certainly be impossible to demonstrate that everything red must have a degree of vividness. Separation of Thirdness, or Tertial Separation, called *discrimination*, consists in representing one of the two separands without representing the other. If A can be prescinded from, i.e. supposed without, B, then B can, at least, be discriminated from A. (Peirce 1998, 270).

Furthermore, these distinguishing abilities are what make the very separation of Peirce's basic categories possible. None of them may be dissociated; but:

13. The central quote is the following:
> The terms "precision" and "abstraction," which were formerly applied to every kind of separation, are now limited, not merely to mental separation, but to that which arises from *attention to* one element and *neglect of* the other. Exclusive attention consists in a definite concepton or *supposition* of one part of an object, without any supposition of the other. *Abstraction* or precision ought to be carefully distinguished from two other modes of mental separation, which may be termed *discrimination* and *dissociation*. Discrimination has to do merely with the senses of the terms, and only draws a distinction in meaning. Dissociation is that separation which, in the absence of a constant association, is permitted by the law of association of images. It is the consciousness of one thing, without the necessary simultaneous consciousness of the other. Abstraction or precision, therefore, supposes a greater separation than discrimination, but a less separation than dissociation. Thus I can discriminate red from blue, space from color, and color from space, but not red from color. I can prescind red from blue, and space from color (as is manifest from the fact that I actually believe there is an uncolored space between my face and the wall); but I cannot prescind color from space, nor red from color. I can dissociate red from blue, but not space from color, color from space, nor red from color.
>
> Precision is not a reciprocal process. It is frequently the case, that, while A cannot be prescinded from B, B can be prescinded from A. (...)

> It is possible to prescind Firstness from Secondness. We can suppose a being whose whole life consists in one unvarying feeling of redness. But it is impossible to prescind Secondness from Firstness. For to suppose two things is to suppose two units; and however colorless and indefinite an object may be, it is something and therein has Firstness, even if it has nothing recognizable as a quality. Everything must have some non-relative element; and this is its Firstness. So likewise it is possible to prescind Secondness from Thirdness. But Thirdness without Secondness would be absurd." (ibid.)[14]

This implies that the three categories are interrelated as follows:

1. ←/→ 2. 2. ←/→ 3.

The categories may not be dissociated.

1. ⟵ 2. 1. —/→ 2.
2. ⟵ 3. 2. —/→ 3.
1. ⟵ 3. 1. —/→ 3.

A lower category may be prescinded from a higher, but not vice versa.

1. ⟵ 2. 1. ⟶ 2.
2. ⟵ 3. 2. ⟶ 3.
1. ⟵ 3. 1. ⟶ 3.

All categories may be discriminated from the others.

This makes the definition of the categories depend on a calculus very close to the mereology of Husserl's 3rd Logical Investigation[15]. The three separation modes may be rephrased as 1) the distinction between autonomous (genuine) parts, 2) the distinction separating a founding content from a founded content, and 3) the distinction separating any moment (founded content, or *unechter Teil*) from its foundational basis. Thus, as in

14. We may note that in this argumentation, the three separation modes are tied to three different modes of presentation: imagining, supposing, and representation, respectively.
15. The relation between Peirce's intense use of the word "phenomenology" in the first years of the century, and Husserl's use of this term is unclear. Peirce refers to the book several times, but his comments on it (as yet another piece of German psychologism) makes it highly improbable that he has in fact read it. It is all the more striking to notice that Peirce's definition of the distinction abstraction types connects them intimately to a part-whole dependency calculus. This idea is exactly parallel to the connection between Husserl's 2nd and 3rd Investigations where the anti-empiricist abstraction theory of the 2nd (abstraction is not inductive generalization, abstraction is a special idealizing focussing act related to an object's properties) and the mereology of the 3rd (the properties thus grasped should be seen as different parts and moments of the object, and a calculus is possible to map these parts' internal relationships). This connection between abstraction and mereology is a highly original idea crucial for the possibility of a realist understanding of the cognition of abstract objects (see Smith (forthcoming), Stjernfelt (forthcoming)).

Husserl, the separation modes are crucial to the explanation of the status of properties (as moments), and their foundation relation (the fact that color properties are founded on spatial properties but not vice versa) and the modes of separation can be seen as the devices necessary for isolating general moments in the phenomenon. Prescission is the most significant type of discrimination because it entails the possibility of isolating properties by leaving other properties in an object indeterminate[16] (corresponding exactly to Husserl's eidetic variation, which inserts algebraic variables in an object for the properties not considered, cf. the Prolegomena to the *Logical Investigations*).

With regard to the semiotic man-animal problem, now, we have still not made much progress. For it must be admitted that many higher animals can perform corresponding acts, as is evident from Kanzi's ability to understand predicate symbols. So let us turn to Peirce's theory of the special kind of symbol called hypostatic abstraction. Peirce often calls prescission "prescissive abstraction" to distinguish it from abstraction proper, or as he calls it, "hypostatic" or "subjectal" abstraction.

While the separation types makes possible generalization – by the peeling away of still further properties – and thus are tied to the Aristotelian general/specific/particular triad, the other abstraction type is tied to the abstract/concrete dichotomy. While the first one is objective – in so far as it discerns objective aspects of the phenomenon – the second is subjective in so far as it is tied to epistemology and to the anatomy of the process of reason (but on this see below).

Hypostatic abstraction is linguistically defined as the process of making a noun out of an adjective; logically as making a subject out of a predicate. The distinction between "hard" and "hardness" serves as the prototypical example. The idea here is that in order to investigate a predicate – which other predicates it is connected to, which conditions it is subjected to, in short to test its possible consequences using Peirce's famous pragmatic maxim – it is necessary to posit it as a subject for investigation. This is evidently a completely different procedure than the separation types (even if the two very often occur interlinked in the research process) insofar as the output is not more general than the input. It makes a second-order object out of a predicate, an object which may now be taken as the object for a further investigation. It takes a thought as a thing, in short. Consequently, the operation is recursive, and we may produce an unlimited hierarchy of ever more abstract notions[17].

In the beginning of the century, Peirce over and over again illustrates abstraction with reference to the well-known Molière joke about the "*Virtus Dormitiva*", the dormitive powers, of opium[18]. The joke is a parody of sterile abstractions of Scholastic medicine, of course, and in Peirce's positivist time, it apparently functioned as a general warning against abstractions *tout court*. But Peirce turns the table on this interpretation. He admits, of course, that it serves as an extreme example of an idle and useless abstraction, but still there remains, if we put it under a microscope, as he says, an ever so small step forward in the reasoning process, even in this foolish example. By going from the statement that "opium puts people to sleep" and to the statement, that "opium possesses a *virtus dormativa*", a hypostatic abstraction has been performed. *Something* in opium is taken to have this effect. We know nothing more positively about the workings of opium in the brain, but the hypostatic abstraction now permits us to ask further: in what, more precisely, does this *virtus dormitiva* consist? It might be that opium just put some people to sleep by coincidence, but the hypostatic abstraction – by substantivizing this ability – asks the question of possible further reasons and structures behind this mere fact. Thus, hypostatic abstraction is a crucial motor in the process of research by positing new *somethings*, new *x*'s, as questions to be investigated.[19] It consists in "asserting that a given sign is applicable instead of merely applying it", as Peirce says already in 1898.[20] It consists in going from saying that something is red to referring to the fact that redness may be applied to something, and in doing so, it creates an *ens rationis*, a second intention, whose truth resides in the fact that something holds for other, really existing things: "For what is an abstraction but an object whose being consists in facts about other things?" ("Logic of History" (1904), NEM IV: 11) The point of the Molière joke, consequently, is not that hypostatic abstraction is futile, but rather that the idea of taking such an abstraction as a *sufficient* explanation is foolish.

In a 1905 manuscript "Basis of Pragmatism" (284), Peirce attempts to give hypostatic abstractions a systematic place in his semiotic architecture. In a chapter on the "Division of Signs", he presents a new trichotomy pertaining to the sign's relation to its immeadiate

16. "In general, prescission is always accomplished by imagining ourselves in situations in which certain elements of fact cannot be ascertained." (2.428 – "Supplement", 1893) Prescission thus is Peirce's version of Duns Scotus famed "formal distinction"; it refers to a distinction made by the mind, but with a *fundamentum in re*.

17. A simple example is the train of thought as follows: *a white particular thing* – (P) – *white things as such* – (A) – *whiteness*, with P for prescission and A for abstraction. A more complicated example is hinted at in the theory of sets and may be reconstructed as follows: *elements* - (P) - *belonging together* – (A) – *a set* – (P) – *bigger/smaller* – (A) – *multitude* – (P) – *relation to other multitudes* – (A) – *cardinal number* ... This example is reconstructed from "Consequences of Critical Common-Sensism" (1905), CP 5.534. The unlimited character of this abstraction process does not entail it is infinite.

18. So as for instance ("Consequences of Critical Common-Sensism" (1905), CP 5.534). The famous quote stems from the third interlude in Molière's last play, "Le malade imaginaire" which introduces a grotesque ceremony of doctors dancing and singing medical latin. Here, a medicine student answers a doctor's question as follows:

> BACHELIERUS/ Mihi a docto doctere/ Domandatur causam et rationem quare/ Opium facit dormire./ A quoi respondeo,/ Quia est in eo/ Virtus dormitiva,/ Cujus est natura/ Sensus assoupire. CHORUS/ Bene, bene, bene, bene respondere./ Dignus, dignus est entrare/ In nostro docto corpore." (Molière, p. 660); in my translation: "Bachelor/ Me the learned doctor/ asks about the cause and reason why/ Opium puts to sleep./ To this I answer/ That there is in it/ A sleep-inducing power/ Whose nature it is/ To weaken the senses.// CHORUS/ Good, good, good, good answer./ Honorable, honorable is it to enter/ into our learned society."

19. Peirce's theory of hypostatic abstraction thus forms a strong argument against the current fad in rhetorics where it is claimed that abstract noun use is just a showoff strategy trying to impress the reader with difficult wording (the opium argument of our time), while texts which express "the same" in more concrete terms are praised as more honest and easier to read. In a Danish context, the two writing styles are even connected to males and females, respectively, rhetoricians taking the party of the latter. If this claim were true, it would do the feminist cause a questionable service as abstractions are indeed necessary for thought to occur.

20. Roberts, p. 64.

object (in contradistinction to the well-known icon-index-symbol pertaining to its relation to its dynamic object). Here, he distinguishes *vague, or indefinite* signs, *singular* signs, and *general* signs, respectively. The singular sign refers to one individual object, while the vague signs refers to objects which need more precise description, and the general signs refer to a possible continuity of objects, among which the interpreter is free to chose as he likes. Among the singular signs, now, a further sub-trichotomy is posited as follows: *hypostatically abstract* signs, *concrete* signs, and *collective* signs, respectively. The last two refer to singular existing things and things built from parts or elements[21], respectively, while the former are characterised thus: "The Immediate Object, though Singular in form, is represented as having the logically material character of the Priman, which is the absence of the matter of existence." (67) As soon as an abstraction is performed linguistically, though, it becomes a symbol, so we should recognize hypostatic abstractions as a specific subtype of symbols. Of course it is possible to refer to abstractions by other means than language – diagrams will be a typical way of referring to them, as we shall see below – but in these cases a symbolic, general indication of their object will be a part of the sign. Collections, the third subtype, are of course already themselves abstractions, and an interesting fact is that also the single existing object for Peirce is an abstraction[22] – it is only possible as a limit case for investigation; in so far it is no wonder that hypostatic abstractions are seen as the most simple singular signs.

Hypostatic abstraction is supposed to play a crucial role in the reasoning process for several reasons. The first is that, by making a thing out of a thought, it facilitates the possibility for thought to reflect critically upon the distinctions with which it operates, to control them, reshape them, combine them.[23] Thought becomes emancipated from the prison of the given, in which abstract properties only exist as Husserlian moments, and even if prescission may isolate those moments and induction may propose regularities between them (and even if we have reason to believe higher animals may perform these two mental

21. Peirce has a basic mereological intuition in so far he refuses to distinguish between parthood and elementhood (as in set theory) and sees those two as shadings of one and the same basic relation.

22. Cf. e.g. "... I do not think that we need have any further scruple in admitting that abstractions may be real, – indeed, a good deal less open to suspicion of fiction than are the primary substances." (Pragmatism Lectures (1903), Peirce 1997, p. 136. This idea forms a very important phenomenological principle in Peirce: the objects we have directly access to are neither completely abstract nor concrete; they are at different intermediate levels. Thus, both abstract objects and the concrete object are constructions reached by abstraction, and the ladder of levels is virtually bottomless; we have no guarantee that it terminates "downwards" in some elementary, atom-like entities. As collections are also abstractions, this consideration also goes for scalar properties. This idea is basically a mereological idea (whole-part relations are pertinent on all levels of observation or reflection) and fits nicely with Peirce's proto-mereological refusal of distinguishing element-of and part-of relations (like it was later formalized in Lesniewski).

23. T.L. Short (1983) traces the idea's roots in Peirce's thought to the famous "How to Make our Ideas Clear" paper from 1878 where Short emphazises the crucial idea that "we *can* use ideas that are *less* clear to make other ideas *more* clear" (290). Thus, the fact that the higher, more abstract terms may be more clear than their concrete basis is a crucial insight in order to avoid infinite regresses and appreciate the role played by abstraction in Peirce.

operations), the road for thought to the possible establishment of relations between abstracta is barred. The object created by a hypostatic abstraction is a thing, but it is of course no actually existing thing, rather it is a scholastic *ens rationis*, it is a figment of thought. It is a thought about a thought – but this does not, in Peirce's realism, imply that it is necessarily fictitious. In many cases it may be – as when we make the abstraction of unicornicity – but in other cases we may hit upon an abstraction having real existence:

> Putting aside precisive abstraction altogether, it is necessary to consider a little what is meant by saying that the product of subjectal abstraction is a creation of thought. (...) That the abstract subject is an *ens rationis*, or creation of thought does not mean that it is a fiction. The popular ridicule of it is one of the manifestations of that stoical (and Epicurean, but more marked in stoicism) doctrine that existence is the only mode of being which came in shortly before Descartes, in consequence of the disgust and resentment which progressive minds felt for the Dunces, or Scotists. If one thinks of it, a *possibility* is a far more important fact than any *actuality* can be. (...) An abstraction is a creation of thought; but the real fact which is important in this connection is not that actual thinking has caused the predicate to be converted into a subject, but that this is *possible*. The abstraction, in any important sense, is not an actual thought but a general type to which thought may conform.
> (Letter to E. H. Moore, Jan. 2. 1904; NEM III/2 918).

The pragmatic maxim reads as follows: if we take all possible effects we can conceive an object to have, then our conception of those effects is identical with our conception of that object. This maxim, with its skeptical implications, never ceases to surprise. But if we can conceive of abstract properties of the objects to have effects, then they are part of our conception of it, and hence they must possess reality as well (see the 1903 Lectures on Pragmatism, Peirce 1997, p. 134). An abstraction is a possible way for an object to behave – and if certain objects conform to this behavior, then that abstraction is real; it is a "real possibility", or a general object. If not, it may still retain its character of possibility (just as Husserl in the Prolegomena to the *Logical Investigations* states that the law of gravity would not cease to hold even if the last heavy object in the universe vanished). Peirce's definitions of hypostatic abstractions occasionally confuse this point. When he claims that "An abstraction is a substance whose being consists in the truth of some proposition concerning a more primary substance." (Peirce 1997, 135), then the abstraction's existence depends on the truth of some claim concerning a less abstract substance. But if the less abstract substance in question does not exist, and the claim in question consequently will be meaningless or false, then the abstraction will – following the definition – cease to exist. But "unicornicity" does not stop being an abstraction just because no unicorns exist. The problem is only that Peirce does not sufficiently clearly distinguish between the really existing substances which abstractive expressions may refer to, on the one hand, and those expressions themselves, on the other. It is the same confusion which may make one able Peirce scholar claim that hypostatic abstraction is a deduction and another – no less able – claim it is an abduction.[24] The first case corresponds to there actu-

24. Helmut Pape: "... the 'abstract in concrete form' brought about by a 'realistic hypostatization of

ally existing a thing with the quality abstracted, and where we consequently may expect the existence of a rational explanation for the quality, and, correlatively, an abstract substance corresponding to the supposed *ens rationis* – the second case corresponds to the case – or the phase – where no such rational explanation and corresponding abstract substance has yet been verified. It is of course always possible to make an abstraction symbol, given any predicate – whether that abstraction corresponds to any real possibility is then an issue for further investigation. And Peirce's scientific realism makes him demand that the connections to actual reality of any abstraction should always be estimated: "every kind of proposition is either meaningless or has a Real Secondness as its object. This is a fact that every reader of philosophy should carefully bear in mind, translating every abstractly expressed proposition into its precise meaning in reference to an individual experience." ("Syllabus" (1903), CP 2.315). This warning is directed, of course, towards empirical abstractions. But in any case the step of hypostatic abstraction is necessary for the ongoing investigation, be it in pure or empirical cases.

The pure case corresponds to the second reason for abstraction's central role in reasoning: hypostatic abstraction's role as a most central operation in mathematics: it is the possibility of making an operation the object of a new operation so as to investigate the rules holding for the first operation (its transitivity, symmetry, etc.). As elsewhere, this abstraction procedure is recursive and may form a hierarchy of concepts. Generalization undertaken by prescission is, of course, equally important in mathematics. Mathematics is linked to hypothetical deduction and diagrams in a very tight fashion in Peirce: mathematics is the science that draws necessary conclusions, diagrams are the vehicles for all deductive reasoning. Deductive reasoning featuring empirical matter must thus imply a diagrammatic, mathematical structure, and diagrammatic reasoning forms a center of Peirce's epistemology: the iconicity of the diagram ensures its structural similarity with its object, the symbol governing it determines the possibility of manipulating it with regard to gaining new information.[25] Both abstraction types play a crucial role in diagrammatical reasoning:

> All necessary reasoning without exception is diagrammatic. That is, we construct an icon of our hypothetical state of things and proceed to observe it. This observation leads us to suspect that something is true, which we may or may not be able to formulate with precision, and we proceed to inquire whether it is true or not. For this purpose it is necessary to form a plan of investigation and this is the most difficult part of the whole operation. We not only have to select the features of the diagram which it will be pertinent to pay attention to, but it is also of great importance to return again and again to certain features. Otherwise, although our conclusions may be correct, they will not be the particular conclusions at which we are aiming. But the greatest point of art consists in the introduction of suitable *abstractions*. By this I mean such a transformation of our diagrams that characters of one diagram may appear in another as things. A familiar example is where in analysis we treat operations as themselves the subject of operations. (Peirce 1997, p. 226)

Thus, the two abstraction types are seminal for diagram formation. Prescission permits us to construct a general diagram, bracketing all contingent features of the particular diagram drawing in favor of the features of it to be read as referring to a general object. Only the required predicates are preserved by this prescission procedure. Abstraction allows diagrams to be recursive and to investigate the properties of other diagrams. Through these two operations, diagrammatic reasoning performed by rule-bound experimentation on the diagrams is made possible. But it is important to notice in our context that we have no reason to suppose that animals may not make simple diagrams (the Y of the fork in the road), nor experiment upon them (the dog's probable inference taking the more travelled of the roads is such an experiment). We have, however, no reason at all for believing that the abstractive making a diagram explicit is a part of higher animals' reasoning abilities.

Seen from a logical point of view, the character of class formation possessed by abstraction makes it belong to the level of what is now called second-order logic (which is, of course, unlimited – in contradistinction to first-order predicate logic, which allows quantification only over individual variables). T. L. Short remarks (1997, p. 295) that hypostatic abstraction is identical with "the transition from first- to second-order predicate logic", and he adds: "It does not follow that every fact about an *ens rationis* is inferable from facts about other things. Second-order predicate logic is not reducible to first-order predicate logic; mathematics could not be done without referring to classes or to other abstract entities." (p. 296) This makes explicit the purpose of abstractions: they are not only shorthand for information already available at the concrete levels. They may add genuinely new information – exactly in accordance with Peirce's idea that, by theorematic reasoning with diagrams (as opposed to merely reasoning through corollaries), new information may appear that was not explicit in the construction of the diagrams at the start.

A more detailed investigation of hypostatic abstraction must try to analyse its basic subtypes. There are obviously many different dimensions along which hypostatic abstractions can be performed. They may give rise to a linguistic variety of semantically different abstract noun types and, more broadly, nominal constructions. I have found no Peirce scholar trying to go this way, but in the Husserlian tradition of pure a priori grammar a scholar like Jean-Louis Gardies has some ideas on linguistic hypostatic abstraction types. There are, for instance, at least three types of possible quotation-marks (in a wide acceptance of the term), each of them nominalizing the expression in question:

1. the operator "the fact that ..." which forms the name of a state of affairs;

relations' is a deductively valid form of reasoning which he at other places calls 'hypostatic abstraction' and which is now called class abstraction." (Pape 1997, p. 171) Pape explains in a note: "It is obviously deductively valid to conclude that, if there is a red rose, the class of red things has at least one member, namely, this rose." (182n). In the very same volume we read T.L.Short: "... neither is the inference to it logically necessary. Rather, that inference could be deductively valid only with the additional, logically contingent premiss that the regularity in question has an explanation (...) Absent that assumption, the inference is not deductive but an extreme case of what Peirce called 'abduction' ..." (297).

25. I have argued for the central role of diagrams in epistemology in Stjernfelt (forthcoming).

2. the nominalization of a predicate ("redness", "humanity");
3. ordinary quotation marks referring to the name of a proposition (or any other element of discourse either in the realm of expression ("or" is pronounced parallel to "door"), or in that of structure ("or" is a conjunction), or in that of content ("or" may mean XOR or it may mean V).

We may add – from Peirce's ideas above:

4. the collection operator forming a set of objects ("my books", "mankind"),
5. the individual object operator: "that object as it exists now and here" (or with any other spatiotemporal or other specification), cf. Peirce's contention that the unique object with all properties completely determined is also an abstract idea.

From traditional linguistics we derive a whole series of cases:

6. verbal substantives permits to abstract a verbal predicate in other ways than ordinary nominalization: present perfect ("operating") forming the abstract idea of an ongoing process; past perfect ("operated") forming the abstract idea of a process having taken place; infinitive ("operate") forming the abstract idea of the process content apart from realization; nominalization ("operation") forming the abstraction of the process as a whole; nominalization of the agent ("operator") forming the abstract idea of a specific ergative subject for a process; adjectivalisation ("operational") forming the idea of some other x having to do with the process. From predicate relations with more than one relative, several different roles may be abstracted (from "give": "the giver", "the gift", "the given"). Other languages may add still further types (gerundive: "the one that ought to be given something" etc.).

But hypostatic abstraction need not be expressed in nor refer to linguistic entities (even if they support it and enhance the possibilities for using it). A recurring example in Peirce is the idea of seeing the geometrical line as an abstraction from the trajectory of a particle. This implies that the nominalization act of hypostatic abstraction also may include the spatial "stiffening" of temporal processes or aspects thereof into objects of an abstract space. All abstraction types probably refer – explicitly or implicitly – to such spaces in which diagrams may take other diagrams as their objects. The description of hypostatic abstraction in terms of linguistic or logical vocabulary should not keep us from finding the phenomenological basis for it, and the possibility for diagrams of taking other diagrams as their objects (thoughts taken as things) precisely presupposes abstract spaces embedded in other abstract spaces. The list of how this may be achieved and represented is possibly open-ended, given the fact that an abstraction of a given predicate may be attempted with reference to any other already performed abstract idea; this open-endedness corresponds to abstraction's homology with second-order logic.

In spite of the fragmentary treatment of these two abstraction types, they play, as is evident, a central role in Peirce's architectonic. In his Carnegie application from 1902, e.g., one of the few occasions when he proposes a systematic exposition of his mature thought, the hypostatic abstraction appears already in lecture 4 (out of 36, and long before the introduction of categories, signs, etc.). This is of course because of hypostatic abstraction's central role in mathematics – the possibility of an operation to be taken as an object for another operation, investigating the first operation's properties. And prescissive abstraction is logically prerequisite to hypostatic abstraction: before hypostatic abstraction of a predicate to a subject, a predicate must already be prescinded. This interplay between the abstraction types are rarely treated explicitly in Peirce, but in one significant passage he links the two with the animal-man transition problem, and this passage is worth quoting at length (in the letter to E. H. Moore, Jan. 2. 1904):

> There are two entirely different things that are often confused from no cause that I can see except that the words *abstract* and *abstraction* are applied to both. One is [*aphaeresis*], leaving something out of account in order to attend to something else. That is *precisive abstraction*. The other consists in making a subject out of a predicate. Instead of saying, Opium puts people to sleep, you say it has a dormitive virtue. This is an all important proceeding in mathematics. For example take all "symbolic" methods, in which operations are operated upon. That may be called *subjectal abstraction*. This use of the word abstract goes back to the beginning of the XIIIth Century while the other use is earlier still. So both are of unquestionable respectability. But they have nothing in common. What I say in treating such subjects I am apt to mean. They have nothing in common. No doubt subjectal abstraction presupposes a certain considerable precisive abstraction in each case; but that was not introduced in making the subjectal abstraction, it was there before. Experience is first forced upon us in the form of a flow of images. Thereupon thought makes certain assertions. It professes to pick the image into pieces and to detect in it certain characters. This is not literally true. The image has no parts, least of all predicates. Thus predication involves precisive abstraction. Precisive abstraction creates predicates. Subjectal abstraction creates subjects. Both predicates and subjects are creations of thought. But this is hardly more than a phrase; for *creation* and *thought* have different meanings as applied to the two. Without precisive abstraction man would not be man; but I can well believe, – indeed, I do think it probable, – that a large fraction of the races of mankind, by no means necessarily very low in the arts, are entirely devoid of the power of subjectal abstraction. (NEM III/2, 917–18).

Lots of interesting ideas are implied in this. Here we find the idea that prescissive abstraction precedes hypostatic abstraction, that the former creates predicates and that the latter, in turn, creates subjects. In so far as even simple collections are abstract entities, it follows that this creation process goes on in human thought all of the time and not only in its purified form in the sciences. Everyday reflection is impossible without it; T. L. Short even argues that the self – and correlatively self-consciousness – is an entity inferred by means of hypostatic abstraction from faults in single actions (as the source of those errors).[26] In relation to the semiotic question of the animal-man transition we find a passing reflection upon the relation of abstraction to biology: the idea that without prescissive abstraction man would not be man, while many human beings, maybe even cultures,[27] may

26. This should, of course, be taken to refer to reflective self-consciousness. Pre-reflective self-consciousness (cf. Zahavi 2000) as a moment of any conscious experience is presupposed by reflective self-consciousness, and may, unlike the latter, probably be found in large parts of the animal kingdom and maybe even in lower organisms.
27. The quote talks about "races", but we should not take this as an indication that any idea of "rac-

lack sufficient ability to perform hypostatic abstractions. This is not, it must be admitted, very precise, and we have already assumed prescission to be widespread in higher animals, as is evident in their ability to associate via qualities (Peirce: "The most ordinary fact of perception, such as 'it is light', involves precisive abstraction, or prescission" (4.235)). In any case, if many higher animals may prescind and man not be man without it, hypostatic abstraction seems restricted to mankind, even if it is perhaps unevenly distributed among us (which might in fact be an indication that selection pressure for it is still at work, or has been until recently).

As is evident from the above, this conforms with our general idea: it is the ability to form not symbols in general, but the special symbol type called hypostatic abstractions, that distinguishes man from (most) animals. It is, of course, a very difficult problem to ascertain which mental procedures higher animals are capable of. But it seems reasonable to assume that they master symbols, including arguments, action according to diagrams, and even symbol systems in some rudimentary form, involving huge amounts of generality made possible by prescission – but with no means to extract that generality from sensory experience and to isolate it, control it or experiment upon it. Here, Peirce's 1905 reflection briefly quoted at the beginning of this paper adds the control dimension as an important role for the abstractions to play:

> Pragmaticist. To my thinking that faculty [of language] is itself a phenomenon of self-control. For thinking is a kind of conduct, and is itself controllable, as everybody knows. Now the intellektual control of thinking takes place by thinking about thoughts [cf. the description of hypostatic abstraction in such second intention terms]. All thinking is by signs; and the brutes use signs. But they perhaps rarely think of them as signs. To do so is manifestly a second step in the use of language. Brutes use language, and seem to exercise some little control over it. But they certainly do not carry this control to anything like the same grade that we do. They do not criticize their thought logically. (5.534)

Man as well as animals are consequently rational beings, probably even necessarily so. Both are involved in a constant series of arguments, in a reasoning process involving a whole range of simpler sign types. But what enables man to build up his symbol systems and its resulting more acute and accelerated rationality is prescission and abstraction working together, making it possible to isolate and to make explicit single phases in the ongoing chain of arguments in order to control them, scrutinize them, experiment upon them, combine them, recombine them, and make them better. Animals may possess the same abilities in germ, prescission probably especially so, but the continuum going from animal to man is to be grasped in terms of gradually higher mastering of abstraction. It must also be admitted, though, that this ability greatly enhances man's ability to commit errors, to be fooled, to lie. Of course, higher animals possess all these abilities, but abstraction adds the possibility for the construction of the enormous subdomains of discourse: myth, religion, literature, science whose capacity for general truths mirrors an equally large capacity for general fallacies.

ism" could be found in Peirce's thought; he is merely using the word as coextensive with "culture", such as was a commonplace at the time.

This would also conform well to Deacon's Baldwinian assumption: that the behavior-selection feedback in symbol-using Homo Habilis communities acquired an extreme pace when measured against evolution's normal velocity. For the active controlling and experimenting on signs makes it possible to develop them significantly within one single biological generation, while the spontaneous historical aspects of language evolution (change in phonetic patterns etc.) is a much slower phenomenon, even if still quick in comparison with biological evolution. Given this scenario, it seems reasonable to assume that a very strong selection pressure has prevailed against the increased possibility of fallacies, especially against those formal logical fallacies which involve no empirical content, but also against violations of basic linguistic constants like the subject-predicate structure.[28] All in all, these abstraction operations permit us to construct an indefinite panoply of abstract objects, more or less apart from the actually surrounding world of here and now – and it permits us, by the same token, to construct diagrams to bring these abstracta to the test, yielding to an already amazing degree increasing insight in empirical regularities as well as in formal and synthetical a priori laws.[29]

If this idea is correct, human beings are a symbolic species, but not the only one. Man is, rather, the abstracting animal.

Literature

Baldwin, J.M.: *Development and Evolution*, New York: Macmillan Company 1902.

Cowan, G.A., David Pines, and David Meltzer: *Complexity. Metaphors, Models, and Reality*, Santa Fe Institute Studies in the Sciences of Complexity Proceedings vol XIX, Reading, Mass.: Addison-Wesley 1994.

Deacon, Terrence: *The Symbolic Species*, New York 1997: W.W. Norton.

— "The Trouble With Memes (and What to Do About It)", (unpublished working paper, Boston 1999).

Emmeche, Cl. and Jesper Hoffmeyer: "Code-Duality and the Semiotics of Nature", in Anderson and Merrell (eds.) *Semiotic Modeling*, Berlin: Gruyter 1992.

Gardiès, Jean-Louis: *Rational Grammar*, München: Philosophia 1985.

Godfrey-Smith, Peter: *Complexity and the Function of Mind in Nature*, Cambridge: Cambridge University Press 1998.

28. The Husserlian idea of a pure grammar – to some extent shared by Peirce - may make the Chomsky-Deacon conflict around grammatical inneism irrelevant. If there are a priori rules for grammar, then we should expect evolution (biological as well as linguistic evolution) to conform to them in a gradual approximation, making the riddle of possible innate chunks of universal grammar easily understandable, because no empirical selection pressure will be needed for their articulation. This would correspond to the fact that we have learned elementary arithmetics and perform that easily without anybody wondering about the specific selection pressures giving rise to an "arithmetic module" in the brain.

29. Thus the idea fits nicely with Barry Smith's insistence on a "fallibilistic apriorism", claiming that a priori laws' validity does not depend on us, but that the *discovery* of them does.

Hintikka, J.: "The Place of C.S. Peirce in the History of Logical Theory", in Brunning and Forster (eds.) *The Rule of Reason*, Toronto: University of Toronto Press 1997.
Hoffmeyer, J.: *Signs of Meaning in the Universe*, Bloomington: Indiana University Press 1996.
Houser, N.: "Peirce as a Logician", in Houser, Roberts, and Van Evra (eds.) *Studies in the Logic of Charles Sanders Peirce*, Bloomington: Indiana University Press 1997.
Husserl, E.: *Logische Untersuchungen*, I, *Hua* XVIII, Den Haag: Martinus Nijhoff, 1975.
— *Logische Untersuchungen* II, I.–II. Teil (Text nach *Hua* XIX/1–2), Hamburg: Felix Meiner 1984.
— *Ding und Raum, Hua* XVI, Den Haag: Martinus Nijhoff 1973.
Molière, J. P. de: *Œuvres complètes*, Paris 1962: Seuil.
Pape, H.: "The Logical Structures of Idealism. C. S. Peirce's Search for a Logic of Mental Processes", in Brunning and Forster (eds.) *The Rule of Reason*, Toronto: University of Toronto Press 1997.
Peirce, C.: *Collected Papers*, I-VIII, London: Thoemmes Press 1998 (1931–58) (references by vol. + Paragraph numbers (as CP 2.130)).
— *New Elements of Mathematics*, (ed. C. Eisele) I-IV, The Hague: Mouton 1976.
— *Pragmatism as a Principle*, (ed. A. Turrisi), Albany: SUNY Press 1997.
— *The Essential Peirce*, vol II, Bloomington: Indiana UP 1998.
Manuscripts from Peirce's unpublished papers (quoted with permission from Dept. of Philosophy and the Houghton Library, Harvard University), numbers referring to Robin 1967.
Prodi, G.: "Signs and Codes in Immunology", in Sercarz et al. (eds.) *The Semiotics of Cellular Communication in the Immune System*, Berlin 1988.
Robin, R.: *Annotated Catalogue of the Papers of Charles S. Peirce*, Worcester Mass.: University of Massachusetts Press 1967.
Short, T.L.: "Hypostatic Abstraction in Self-Consciousness", in Brunning and Forster (eds.) *The Rule of Reason*, Toronto: University of Toronto Press 1997.
Smith, B.: *Austrian Philosophy*, Chicago: Open Court 1994.
— "Logic and Formal Ontology" in *Manuscrito* (forthcoming).
Sowa, J.: "Matching Logical Structure to Linguistic Structure", in Houser, Roberts, and Van Evra (eds.) *Studies in the Logic of Charles Sanders Peirce*, Bloomington: Indiana University Press 1997.
Sterelny, Kim, and Paul E. Griffiths: *Sex and Death. An Introduction to the Philosophy of Biology*, Chicago: The University of Chicago Press 1999.
Stjernfelt, F.: "Biosemiotics and Formal Ontology", *Semiotica* 127 – 1/4, 1999, 537–66.
— "How to Learn More. An Apology for a Strong Concept of Iconicity", in T. D. Johansson et al. *Iconicity. A Fundamental Problem in Semiotics*, Copenhagen: NSU Press 1999, 21–58.
— "Diagrams as Centerpiece of a Peircean Epistemology, in *Transactions of the Charles S. Peirce Society* , vol. XXXVI, no. 3 (forthcoming).
— "Mereology and Semiotics", in *Sign Systems Studies*, vol. 28, Tartu 2000 (forthcoming a).
— "Categories, Diagrams, Schemata", in Stjernfelt & Zahavi (eds.) *100 Years of Phenomenology. Logical Investigations Revisited* (forthcoming b).

— "A Natural Symphony? von Uexküll's "Bedeutungslehre" for our days' semiotics", in *Semiotica* (forthcoming c).
Zahavi, Dan: *Self-Awareness and Alterity. A Phenomenological Investigation*, Evanston: Northwestern University Press. 1999.
Zeman, J. J.: "Peirce on Abstraction", in E. Freeman (ed.) *The Relevance of Charles Peirce*, La Salle, Ill.: The Hegeler Institute 1983.

Privileged Rationality

AVRUM STROLL

As Wittgenstein tells us there are many kinds of language games, each of which has its special point and purpose. But it is compatible with this position to assert that there may be a particular game that is superordinate to any other game that purports to have the same point or purpose. I believe that with respect to the investigation of the real world the scientific game is fundamental in this sense. It uses a methodology that defines rationality across all language games in which rational inquiry is a central aim. Whether the discipline is history, sociology, linguistics, anthropology, mathematics, or psychology, it must satisfy these methodological procedures if its pursuits are to be rational. As with any intricate game, scientific methodology is complex and the list of its main tenets is indefinitely long. I will simply list a small handful of these. I will also follow Wittgenstein's advice that one should not think of these provisions as necessary or sufficient conditions, or both. A rational game may satisfy some of them and fail to satisfy others.

First, any problem must be addressed with a minimal degree of illusion, emotion and subjectivity. Second, a rational inquiry should be based on cogent argumentation. Third, such an inquiry should start from whatever evidence is available and should derive its inferences from evidential findings. Fourth, where there is no evidence, or where evidence is indecisive, rationality requires a suspension of judgment about a particular conjecture. Fifth, science posits that many features of the world are mind-independent. Thus, any view that holds the world to be entirely mental is irrational according to this criterion. But since the concept of mind independence entails that minds exist, any view that denies the reality of the mental is also irrational. Sixth, the goals of a rational inquiry are to expand the existing sum of knowledge and to arrive at a true account of reality. It is, of course, possible to aim at knowledge and truth and yet to be mistaken. In such a case the inquiry is not irrational. There are thus cases of rational activity that are simply erroneous. What is important is that if they are rational they aim at knowledge and truth, and employ most of the methodological principles mentioned above.

In the light of these remarks, we may define an inquiry *whose purpose is knowledge or truth* as irrational if it violates a weighted set of these criteria. This definition of "irrationality" does not imply that every conceptual endeavor that fails to satisfy these criteria is irrational. If the aim of a poem is to instill a certain emotion or feeling in a reader its aim is not, generally speaking, to be judged as irrational. It is engaged in an activity that falls outside the domain of rational activity. We can say in such a case that it is not playing the rational game. Thus many intelligent activities are to be characterized as *arational* rather than as irrational. I would say this is generally true of music or poetry and certain kinds of abstract art. It is also possible that a given discipline – philosophy being a notable example – may exhibit features that are both truth-seeking and non-truth seeking, and accordingly can only be judged as rational or as arational depending on the goals and purposes that particular aspect of the discipline is aiming at.

B. Brogaard, B. Smith (Hg.), *Rationality and Irrationality / Rationalität und Irrationalität*, S. 362–369.

In the spirit of the later Wittgenstein, I should also emphasize that there are many different kinds of irrationality, from insanity to obduracy in the face of counter-evidence. In everyday life these cases form a spectrum. We encounter contemporary persons who claim to be the risen Christ, paranoids who have a tightly reasoned but insane view of things, and some individuals who think they will win the California lottery. This last case is more moderate than the first two since some persons do win the lottery – the odds being something like forty million to one – whereas there is no evidence that anyone living today is identical with the historical Jesus. They thus represent different degrees of irrationality. This spectrum from the crazy to the merely implausible is also exhibited by philosophical theories, both past and present. In my view, it is the mildly irrational cases that are the most interesting. They raise the question whether there are criteria of rationality other than the scientific ones I have mentioned that would allow us to distinguish various degrees of philosophical irrationality from one another, and all of these from philosophical views that are so strange as to verge on irrationality without being so. These last I will call cases of *privileged rationality*. I will argue that there are such criteria. What these are I will explain later. My general thesis is that all these cases – from the profoundly irrational through the privileged rational – arise from a common source, and I will also explain later what I take this to be.

Let us make all this more explicit by identifying some clear examples of philosophical irrationality, starting with two extreme cases. Parmenides, while walking with his disciples, denied that motion is possible. Parmenides' theory satisfied the condition that a rational inquiry should be based on cogent argumentation. But it violated the scientific principle that the conclusion arrived at should be based on evidence – "evidence" here meaning "data based on observation." There is no such evidence. Indeed, there is overwhelming evidence against that position. As Wittgenstein says: "Thus, we should not call anybody rational who believed something despite the scientific evidence." (O.C. 324). By this standard, therefore, Parmenides's theory was palpably irrational. Such radical deviations from rationality were, of course, not confined to the Greeks. At the beginning of the last century it was commonplace among British idealists to deny that space and time existed. The fact that these persons traversed space in order to be on time for an appointment entails that their views were also irrational. A less radical form of irrationality is La Mettrie's thesis that human beings are nothing but machines. He cogently argued for this conclusion, and moreover produced some evidence in its favor. As a physician, he noted that bodily processes operate according to mechanical principles: the heart is a pump, the vascular system a set of pipes, and the nerves but so many strings. But his mechanical metaphor quickly lapsed into paradox. Human beings are not devised for specific purposes as are a sewing machine and a renal dialyzer, and they do not have to be set in motion by sentient beings to perform work in the way that machines do. To insist that humans are machines is thus a species of irrationality, but moderate because it is not without evidential support.

Let us now contrast La Mettrie's view with a case of privileged rationality. Such a case is described in Norman Malcolm's essay, "Moore and Wittgenstein on the Sense of 'I Know.'" In *On Certainty* Wittgenstein was so impressed by Malcolm's account that he alludes to it twenty-three times (entries 347, 349, 350, 352, 353, 387, 389, 393, 443, 451, 452,

463, 465, 467, 468, 481, 483, 503, 520, 532, 533, 585, 591). Malcolm writes:

> When we sat in the back garden of his home on Chesterton Road, arguing over the concepts of knowledge and certainty, Moore, wanting to give an example of something he knew for certain would point at a tree a few feet away, and say, with peculiar emphasis, "I *know* that *that's* a *tree*." He would then claim that he had just made an assertion that was perfectly meaningful (as well as true); and I would dispute this claim. (*Thought and Knowledge*, 1977, p. 173).

Here is one of the passages from *On Certainty* in which Wittgenstein describes the preceding scene:

> I am sitting with a philosopher in the garden; he says again and again "I know that that's a tree," pointing to a tree that is near us. Someone else arrives and hears this, and I tell him: "This fellow isn't insane (verrückt). We are only doing philosophy." (O.C. 467).

This passage is interesting for several reasons. The fact that the philosopher obsessively reiterates this remark might well be taken as a sign of dementia. Elsewhere in *On Certainty* Wittgenstein describes a man who is looking for something in a drawer. He opens it, looks into it, and then closes it. He repeats this process interminably. Wittgenstein says that the man has not learned the language game of searching for something. Such distorted behavior is palpably irrational. (Note that by describing the philosopher as saying again and again, "I know that's a tree," Wittgenstein has added a dimension to the scenario that is not in Malcolm's original account).

Wittgenstein's description makes it plain that the tree is close by ("in unsrer Nähe"). It also implies that the philosopher has normal eyesight, that the conditions of visibility are good, and therefore that no ordinary doubt exists that must be resolved. But if all this is so, why does the philosopher repeatedly say "I know that's a tree?" Surely such linguistic behavior is peculiar. And yet, Wittgenstein absolves the philosopher of irrationality. He feels it necessary to assure the unidentified passerby – the proverbial "plain man" – that the philosopher is not a lunatic, and does so by implying that philosophical activity exhibits a special kind of oddity that society should tolerate. Moore is not demented. He is not asserting that he is the risen Christ or that motion is an illusion. In part what makes Moore's remark differ from such genuine cases of irrationality is that there is indeed a tree in front of him. What he says is based on the evidence of his senses and is supported by such evidence. Malcolm and the unidentified passer-by can also see the tree. In contrast to what Moore is saying, there is no observable evidence for the proposition that space, time, and motion do not exist. Wittgenstein is thus justified in pointing out that this is not a case of irrationality. But in stating that "we are only doing philosophy," Wittgenstein is doing more than merely exonorating Moore. He is also implying that there is a privileged species of rationality. It consists of cases, situations, and remarks that would be judged as irrational by ordinary folks and yet are not. They simply satisfy some other conditions of rationality. As we shall see, these are conditions that science would also accept.

What are these conditions? I will mention two. The first is not stated explicitly but is implied by what Wittgenstein says in a number of passages in *On Certainty*, among them 93, 117, and 191. The second is mentioned explicitly in entries 298 and 336. Here is what 117 says.

> Why is it not possible for me to doubt that I have never been on the moon? And how could I try to doubt it?
>
> First and foremost, the supposition that perhaps I have been there would strike me as *idle*. Nothing would follow from it, nothing be explained by it. It would not tie in with anything in my life.
>
> When I say "Nothing speaks for, everything against it," this presupposes a principle of speaking for and against. That is, I must be able to say what *would* speak for it.

In this passage, I take Wittgenstein to be saying that if *nothing* speaks for an hypothesis and if *everything* speaks against it, the hypothesis is irrational in the highest degree. The words "nothing" and "everything" are key here. Nothing would follow from such a conjecture and nothing would be explained by it. The thesis that motion is impossible is such an hypothesis. Nothing speaks for it and everything speaks against it.

With a slight modification, this criterion for extreme irrationality can be turned into a criterion for privileged rationality. The modification is to the effect that if at least *something* speaks for a certain point of view and if *nothing* speaks against it, the concept is doing some work and accordingly is not wholly irrational. The criterion also helps explain why La Mettrie's idea that men are machines is moderately irrational. Something speaks for it, to be sure, but something also speaks against it. Moore's idea that looking at a tree is a paradigm of knowing something, no matter what the circumstances, is not irrational according to this criterion. That a tree exists before him is evidence that speaks in favor of his point of view and no evidence speaks against it. This, then, is a criterion for privileged rationality. There is no evidence that a tree does not exist in his garden, or that Moore is hallucinating, dreaming, or suffering from poor eyesight. What Moore says verges on the irrational without being so. What he says is wrong but for reasons other than lacking evidence or running counter to it.

Let us turn now to our second criterion. In *On Certainty* – as he does elsewhere in his later writings – Wittgenstein emphasizes that what counts as normality is determined by consensual community judgment. In 298 he states:

> 'We are quite sure of it' does not mean just that every single person is certain of it, but that we belong to a community which is bound together by science and education.

The community judges certain views to be irrational no matter how cogently argued they are: such views include paranoia and doctrines to the effect that motion, time and space are unreal. The community also acknowledges that some views are strange without being irrational. Their proponents are accepted by the community; they are regarded as eccentrics, to be tolerated but not taken seriously. Philosophers who espouse theories of privileged rationality fall into this category.

As I previously mentioned, what I find interesting about the whole spectrum of philo-

sophical activity from extreme irrationality through milder forms of irrationality down to cases of privileged rationality, is that they all arise from a similar motivation. That motivation, I submit, is that each philosopher is driven by a particular conceptual model or vision of reality. Such conceits, elaborated in the hands of philosophical masters such as Plato, Hobbes, and Kant, typically exhibit four features. They are paradoxical, homogenize or absorb counter-examples, hold universally with respect to the cases under consideration, and resemble or imitate scientific theorizing. In cases of privileged rationality, as distinct from cases of extreme irrationality, such visions obtain their power from selected examples that are based on observational evidence. The philosopher will commonly argue that these examples have universal scope, i.e., that they are the deepest and most basic cases and that all other examples or cases should be assimilated to them. It is this procedure that gives the model its argumentative force.

The history of philosophy is replete with such examples. Kant's view that the structure of the human mind conditions the objects of perception, and accordingly that such objects are never perceived as they are in themselves, is one example. There is no doubt that on some occasions what we drink or eat conditions what we see. But the generalization from such special cases to all cases is illicit. Illicit, I say, but not irrational since some evidence speaks for it.

Plato's *Republic* is a treasure house of such cases. The notion that the objects of sense experience are subject to change and hence cannot be known is a famous example. The contention that only an intellectual elite is competent to rule society is another. It rests upon the thesis that ruling is a kind of skill. The example used to support this thesis is that only the well-trained physician, the physician *qua* physician, is competent to treat a patient. The notion that the practice of medicine requires the mastery of a body of information and the requisite training in diagnosis is supported by past experience. Those without such skill are not likely to be good physicians. Plato then argues that ruling is a skill analogous to that of medicine. Accordingly, a well run society requires rulers who have mastered that skill and thus have the expertise to rule properly.

But the analogy is faulty and the argument based on it is fallacious. That medicine is a skill cannot be denied. Thus evidence speaks for the analogy. Note that nearly every major university in the world teaches students to acquire such a skill. But also note that there is no university that has a school or department that teaches the skill of ruling. That is because ruling is not a skill in the way that medicine is. Thus what speaks against the Platonic analogy is not counter-evidence; it is community judgment. No community *in fact* operates according to Plato's formula. We thus have here a case of privileged rationality – something speaks for Plato's view and no evidence speaks against it.

The previous examples and my comments on them are perhaps too brief to carry conviction. I therefore propose to deal with a contemporary example of privileged rationality in detail and hope that this will be more compelling. For this purpose I will discuss Hilary Putnam's Twin Earth scenario. Since the scenario is well known I will not describe it further here. But it should be emphasized that it turns on two sets of concepts: identification/misidentification and successful/unsuccessful reference. That is, the scenario entails that an earthling who travels to twin earth will be mistaken in identifying the potable fluid there as water. As Putnam says that liquid cannot be water because water is necessarily H_20 and the fluid on twin earth is not composed of H_20. Thus, "water" as used on earth

does not successfully refer to the substance on twin earth. Putnam's argument in support of his more general thesis about the rigidity of natural kind terms depends on this example. And this example, like others he uses about molydenum and aluminum, and beech and elm trees, is used to make the point that anyone can misidentify a substance unless that person is a trained expert. It is remarkable that, in a wholly different context, Putnam should be echoing the Platonic argument about the need for expertise. Like Plato's, Putnam's argument is both powerful and mistaken; and we can show why it is both by looking at some different cases in which someone or something is identified or misidentified and in which reference may be successful or unsuccessful.

Case I: I am teaching *Naming and Necessity*. I say: "In this work the author, Saul Kripke, has given an ingenious argument whose thrust is that some necessary propositions are also a posteriori." I have identified the author and I have successfully referred to him. We can call this the ideal case where identification and successful reference coincide.

Case II: The year is 1999. The winner of the match is Martina Hingis. Over a loudspeaker, I congratulate her using the name "Martina Navratilova." The crowd groans. They know and indeed I know that Navratilova retired years ago and that the winner is Hingis. I just misspoke. My reference was successful; I congratulated Hingis, the winner. But I misidentified her.

Case III: As a victim of the robbery, I have been asked by the police to pick out a person from a line-up of suspects. I have background reasons for believing the robber is named "Jones." I say: "Jones is the third man from the left. He was the one who did it." Later it is discovered that the third man from the left is not named "Jones," was not even in the country on that date, and that another has confessed to the crime. I misidentified the perpetrator by the name I used and my reference was unsuccessful.

Case IV: It is 1849 and along with a number of other miners I am panning the Sacramento river. Separated from gravel and mud, a huge nugget appears on the mesh of the shallow dish I am using. It glints in the sunlight. I shout "It's gold; I'm rich." Later analysis shows it to be iron pyrites. I misidentified the substance, using the wrong name for it, and my reference, being to gold, was unsuccessful. This example is like Case III except that it is not about an individual person or place but about a natural kind.

Cases III and IV exemplify the kinds of instances mentioned by Putnam in the Twin Earth scenario. They are not like II where reference succeeds even where there is misidentification. In both III and IV there is misidentification and unsuccessful reference.

We are now in a position to state the exact nature of the Twin Earth challenge. In cases III and IV the speaker has a certain object (the thief) or a certain substance (gold) in mind. He misidentifies these items and fails to refer. Putnam's point with respect to proper names is that the relationship between name and object in a referential context is direct and does not depend on the intention of the speaker at all. The name refers because it rigidly designates the object. A similar account holds between natural kind terms and natural kinds such as gold and water. The terms refer because there is a direct relationship between them and the corresponding substance. What a speaker has in mind is thus irrelevant to the referential connection.

What then is wrong with the Putnam scenario? As Wittgenstein tells us, such conceptual models contain genuine insights; but he also emphasizes that by their very generality

they make an accurate apprehension of the world impossible. The Twin Earth scenario is a model of this sort. It creates a picture that induces us to look at the world selectively. It does so by aping scientific theorizing. To see where Putnam has gone wrong, one must start by looking at the model – The Twin Earth scenario – more closely. It will emerge that it creates a disposition to treat diverse examples as if they all fell under the scope of the model. It is a powerful notion but it leads to misapprehension and a misdescription of reality.

Let us concentrate on what the model says about natural kinds, and especially about water. Suppose we ask: "Is it literally true that there could be such a Twin Earth as Putnam suggests?" The answer as I will indicate in a moment, is "no." But if it is "no" then we have a second question to ask: "If that is so, what is the philosophical point being made by using such a scenario?" And the response to this question will be an analysis along the lines I have intimated above, namely that Putnam is using the scenario to impel us to look at the world in a certain way. What that is we shall see below. But now let us address the first question: "Is it true that there could be such a Twin Earth as Putnam suggests?"

According to physicists and chemists whom I have consulted, the Twin Earth scenario is not physically possible. They state that there cannot be two substances that have all properties in common and yet have different microconstituents. As Putnam describes Twin Earth, the substance called "water" there has all the properties of earth-water, except that it is composed of chemicals, X,Y,Z , that are entirely different from hydrogen and oxygen, the components of the substance called "water" on earth. Though this is a logically possible scenario it is not physically or empirically possible. Putnam has the science backwards. There are so-called "geometrical isomers" which are substances having exactly the same components but which have different properties. Ethyl alcohol, whose chemical formula, is C_2H_5OH, and methyl ether, whose chemical formula is CH_3OCH_3, are each composed of two carbon atoms, six hydrogen atoms, and one oxygen atom. But because the atoms in each molecule bind to one another in different ways, the two substances have different properties. Ethyl alcohol is potable for humans and methyl ether is not, for example. So in isomers we do find two different substances having the same constituents. But there are no cases in which differing molecular constituents give rise to two substances having the same properties. In fact, then, Twin Earth is not a possible earth in the empirical sense of "possible." Accordingly, we can discount the Putnam scenario as being literary fiction rather than as scientific fact.

What, then, is its point? I submit that it is designed to induce the reader to look at *every* substance on earth as if each were like gold or iron pyrites. In other words, it is designed to make the reader think that the identification of every substance on earth depends on knowing what its chemical properties are. A further aim of the scenario is to show, following the previous principle, that in referring to any substance we are never referring to its observable properties, but to its microscopic constituents. This is the whole point of the gold example. As we have seen in Case IV, a veteran miner might be deceived into thinking that the piece of iron pyrites in his pan is gold. This is because he is looking at the visible properties of the ore in the pan. Putnam's point is that its overt properties are not sufficient to allow anyone accurately to ascertain what that mineral is. One needs expertise – an assay done in a laboratory by a specially trained professional – to determine whether the ore is gold or fool's gold. The thrust of the argument is to compel his readers to think that *every* substance on earth is like gold or iron pyrites, i.e., that the ordinary person *must* know what its microcomposition is before he or she can accurately identify it. That this is true of many substances is beyond doubt. The average person shown a jar containing the rare earth, Hafnium, would see a gray powder and would not be able to identify it. But not every substance on earth is a rare earth. Some are such that they can be accurately identified by ordinary persons, and water is one of those substances.

Suppose one is looking at two jars filled with transparent liquids. From where one is standing it is impossible to detect any differences between them. But then one is asked to smell or taste each. Differences appear immediately. If one of the substances is pure water it will be odorless and tasteless. If the other is pure ethyl alcohol it will have a distinctive odor and a powerful taste. It does not require an assay to tell the difference. There is nothing on earth that has the same total set of properties that water does. And accordingly, it does not require an assay to distinguish water from some of these other substances. This point can be illustrated in a different way. Since time immemorial human beings have used water. They have drunk it, irrigated plants and trees with it, washed clothes in it, employed it as a solvent, sailed on it, and used it for innumerable other purposes. The discovery that water was composed of hydrogen and oxygen was made only in the middle of the 19th Century. It is clear that people knew what water was long before atomic theory was invented. That knowledge was based on their observation of water, and experience with its overt properties. Thus, without knowing any science they were able to distinguish it from *all* other substances, even though some of those substances were visually similar to it. That they were able to make such discriminations demonstrates that their ability to identify water did not then and does not now depend on any scientific discovery or anything equivalent to a technical assay. Similar remarks apply to animate entities, as well as to such inanimate substances as water. Practically every normal adult can tell the difference between a cow and a sheep, or a dog and a cat. It does not require an examination of the DNA of these differing species to be able to identify specimens of each with complete accuracy.

Since that is so it follows that the Putnam claim that a certain substance can be identified as water only after a chemical analysis is mistaken. What has gone wrong here, then? Like many powerful and insightful philosophical theories the Twin Earth scenario is not without insight. It makes a compelling point with respect to *selected* examples of things, species, and substances. But its mistake is to think that its vision of the world is applicable to *all* persons, species and substances. It thus involves an illicit generalization. This essentially consists in the assimilation of familar natural kinds, such as water and dogs, to natural kinds that require technical expertise for their identification. But as my counter-examples show what holds for a limited number of cases does not hold for all. In flying in the face of scientific evidence Putnam's scenario verges on the irrational. Yet it is not irrational since some evidence speaks for it. I say it is thus a case of privileged rationality.

Multiplicity of Mental Spaces

BARBARA TVERSKY

Erroneous and Precise Spatial Actions

When asked the direction between Phildadelphia and Rome, most people err. They say that Philadelphia is north of Rome when in fact, it is south of Rome. This cannot be dismissed as the weather, because when asked the direction between Boston and Rio, a majority of people erroneously say that Boston is east of Rio. Nor are these errors a simple consequence of randomness, nor of ignorance of geography. Rather, they are systematic and predictable outcomes of the way spatial information is organized in the mind. Contrast these errors with the precise artistry of violin playing or basketball or surgery, or wending one's way through a crowd. How is it that one set of spatial behaviors is predictably erroneous and the other predictably precise? That one set of spatial behaviors appears blunt and irrational and the other delicate and tuned?

How, then, to reconcile spatial behaviors that are finely-tuned and precise with those that are clumsy and erroneous? A closer look at the finely-tuned, precise ones reveals that these behaviors are repeated, indeed practiced, in structured environments that provide feedback and that are replete with cues that guide and support behavior. Sheer repetition in supportive environments provide these perceptual-motor performances the benefits of selection, by evolution or by learning. Behavior that yields good outcomes is selected and repeated; not so behavior whose outcomes are not favorable. Rich environments provide cues for the behaviors more likely to yield favorable outcomes. In short, through practice with feedback, actions become precisely tuned. By constrast, spatial behaviors that are infreqent and based on hypothetical rather than experienced environments, notably judgements, do not enjoy the benefits of either repetition with feedback or of supportive environmental cues. Instead of being situated in real environments, they are mediated by cognitive representations of environments that are local, ad hoc, and schematic. Although the cognitive structures constructed to represent environments and to enable judgements on them are based on qualities of the perceptual world, the fact that they are schematic introduces systematic error and the fact that they are ad hoc and local means there is no guarantee of global coherence, introducing the possibility of inconsistency.

Ad hoc, local, schematic cognitive structures are invoked when there are no means, either by the tools of the mind or by the tools of the world, of producing complete representations on which appropriate and accurate calculations can be performed. Given the constraints of human information processing, especially limited capacity working memory, the use of ad hoc, local, schematic cognitive structures can provide relatively quick and efficient estimations. Some of the errors inevitably introduced by these heuristics may be corrected by constraints in the field or by constraints from other estimates or modes of estimation that may be independent. Other errors, however, may go undetected. Some may lead to unfortunate consequences.

B. Brogaard, B. Smith (Hg.), *Rationality and Irrationality / Rationalität und Irrationalität*, S. 370–378.

What cognitive mechanisms are responsible for spatial behaviors that are precise and tuned and what cognitive mechanisms underly spatial behaviors that are schematic and biased? Precise spatial behaviors, from infants reaching for toys to individuals finding their ways home to couples dancing in consort develop with feedback from extensive practice in specific environments. Spatial skills such as these are situated, that is, executed in environments replete with cues that support, constrain, and guide accurate performance. The specific instrument of the musician, the bodies and tools of the surgeon, the landmarks and layout of the navigator's environment serve to both constrain the spatial behaviors and to cue them. Initial attempts at reaching or wayfinding or dancing are clumsy and fumbling. With practice, the behaviors become accurate perceptual-motor skills that can be enacted automatically, with little demand on attention.

Contrast these cognitive mechanisms with those underlying answering questions or filling requests such as these. "You're standing in the Piazza Navona, facing Bernini's Fountain of the Four Rivers, what's to your left?" "How far is Philadelphia from Pittsburgh?" "From New York City?" "Can you draw me a map to get to your house?." Why is it that such questions are answered from ad hoc, local, schematic mental constructions? As usual in biological systems, there is more than one reason. One factor, already mentioned, yielding schematic mental representations is constraints on working memory. Another is the nature of long-term memory representations. Questions such as those above are hypothetical at the time they are asked; that is, the responder may not be in the environment asked about. Or the questions may be about environments too large to be experienced at once. In order to answer the question, the responder must construct a mental representation of the environment, and then make a mental judgement on the mental representation. Thus the the question must be answered, the judgment must be made on a mental construction of the environment, not on the actual environment. Both the mental construction and the judgement are made in working memory which has a limited capacity (e.g., Baddeley, 1990). Constructing a schematized version of the environment tailormade for the particular query is one way of conserving working memory space.

There is yet another factor encouraging mental constructions of environments that are ad hoc, local, and schematic. Unlike maps people can consult on-line or on the shelves, there may be no ready-made mental representation of the information demanded by the task. This means that the mental representation must be created on the fly from information in long-term memory. The information about environments in long-term memory does not resemble in either structure or format a map such as those developed by cartographers for a variety of purposes. "Cognitive maps" – that is, whatever mental information is used to make geographic judgments – are notoriously inaccurate and incomplete. What's more, they seem to have a variety of formats consisting in part from memory for maps that may have been studied, in part from memory for experiences in environments in part from memory for descriptions of environments. Because of the diversity of formats that have no natural or standard means of integration for the mental representations of spaces, "cognitive collage" has been suggested as a more apt metaphor than "cognitive map" (Tversky, 1993). In order to integrate these different pieces of information, especially when they are in different formats, a common structure is needed. A common structure can be provided in several ways, for example, by landmarks or features common across pieces of information coded by a common reference system. A pictorial memory of

an environment including the city hall and the post office could be combined with a verbal memory of the route instructions from the post office to the bank by the common feature of the post office to allow construction of a route from city hall to the bank Or a memory of a subway map may be integrated with memories of the neighborhoods around subway stops to allow estimates of directions between landmarks at different subway stops. However the bits of information are combined, they are combined roughly and schematically.

As noted, answers to queries about the spatial world are prone to error because the answers are determined in working memory, which has capacity constraints. When answers are not known, they are determined by mentally constructing an ad hoc representation of the space and performing a computation on that representation. However, even if working memory capacity were augmented by using an external representation for computation (e.g., Donald, 1991; Norman, 1993, Tversky, in press), the likelihood of systematic error remains because the sketch maps would be produced schematically combining incomplete fragments of information in different formats. In fact, sketches also demonstrate the sorts of errors and biases that will be described in detail below.

The solution to both the constraints of working memory and the nature of long-term memory representations of environments is mental structures that are tailor-made for the query at hand, constructed from different pieces and kinds of information, selecting only the information relevant to the query, integrating it through a common schematic framework. What's more, the common schematic framework varies with the type of query. Let us now turn to two case studies illustrating this process, one for the space around the body, the other for the space of navigation. Each of these spaces is conceived of differently, in ways that depend on how the space is perceived and used in daily life, that is, on perceptual and functional aspects of the space. Thus, space, as conceived of by ordinary people as they interact with it differs from space as conceived of by geometers or physicists. For the latter, these different spaces are coherent and conceptualized in a unitary fashion.

The analysis of the two spaces will illustrate these features of spatial cognition: that different spaces are conceived of differently, that the way the spaces are conceived depends on perceptual and behavioral experience with them, that the spaces are conceived of schematically, and that the schematization yields bias and distortion.

The Space around the Body

As we move about in the world, we remain aware of the things around us, even the things that are no longer in view. As we move forward, what was formerly ahead of us is now behind us; as we turn, what was formerly ahead of us is now to the side. We are aware of the surrounding scene even when we do not actually perceive it. After all, our range of perception is small, and for the most part, oriented forward. As long as things don't move, we can reach to the side or move backwards without looking and without erring. This awareness of the surrounding situation under movement extends beyond environments acquired by perception and navigation to hypothetical environments acquired only by description (Franklin and Tversky, 1990; also, Bryant, Franklin and Tversky, 1992; Bryant,

Tversky, and Lanca, 2001). Imagined environments and imagined navigation in imagined environments capitalize on the extensive experience people have in navigating real environments, on the mental structures people construct in order to maintain awareness of surroundings. In several dozen experiments, participants studied descriptions of observers (addressed as "you") in environments such as a museum or hotel lobby, surrounded by objects, such as chandeliers and fountains, to their front, back, head, feet, left, or right. They were then told that they have turned to face an object formerly at their sides, back, head, or feet and queried for the objects currently at back, front, left, right, head, and feet. Participants were able to answer these questions quickly and accurately, indicating that people are able to construct mental models of space from descriptions and quickly update their hypothetical postitions in them. Interestingly, although care was taken to make sure that no part of the environment was privileged and that objects were randomly assigned to positions around the body, times to retrieve the objects depended on the direction queried as well as the orientation of the observer, upright or reclining. That is, certain directions from the body were retrieved faster than others for reasons having nothing to do with the particular content of the space or objects in it.

The biases in retrieval times could be accounted for by the Spatial Framework Theory (Franklin and Tversky, 1990). Human bodies have three natural axes, formed by head and feet, front and back, and left and right. The head/feet and front/back axes are asymmetric, but the left/right axis is relatively symmetric. The asymmetries make the poles of the axes more discriminable. The world, too, has three natural axes, only one of which is asymmetric, the up-down axis formed by gravity. When a person is upright, the up/down axis of the world coincides with the head/feet axis of the body, enhancing the discriminabilty of that axis. According to the Spatial Framework Theory, retrievability is enhanced by discriminability so that times to retrieve objects at head and feet should be fastest, because of the confluence of asymmetry of the body and asymmetry of the world, followed by times to front and back, with an asymmetric body axis. Times to retrieve objects at left and right should be slowest as that body axis is for the most part symmetric. This pattern of retrieval times in fact emerged in many studies. The situation of the reclining observer is different. Here, the observer rolls from front to back to side so that no axis of the body is aligned with the asymmetric axis of the world. When just the body is considered, the front/back axis is more salient than the head/feet as the front/back axis separates the world that can be seen and manipulated from the world cannot be readily seen or manipulated. Again, this analysis of asymmetry, hence discriminabilty, conforms to the pattern of retrieval times obtained for the reclining observer: retrieval times to front and back were fastest, followed by times to head/feet and then left/right.

Many variations of these situations have been studied, with consequent variations in the patterns of retrieval times. Some of the situations studied include those where the viewpoint was outside the scene looking on rather than surrounded by objects, those where there were two people each surrounded by a different objects, and those where the room was described as rotating rather than the observer (in order: Bryant, et al., 1992; Franklin, et al., 1992; Tversky, Kim, and Cohen, 1999). The environments have been conveyed by description, by actual experience, by models, and by diagrams (in order: Franklin and Tversky, 1990; Bryant, et al., 2001; Bryant and Tversky, 1999). In all cases where directions or objects were retrieved from memory, the pattern of retrieval times fit the

spatial framework theory, indicating that the memory representations established from experience and from language are functionally similar.

The point of elaborating this example is to illustrate how simple spatial judgements are made in memory. A schematic mental model of the relevant portion of the environment is constructed in working memory, and the judgement made on that representation. The spatial mental model constructed is not arbitrary; rather, it reflects people's conceptions of the spatial world, in this case, the human body and surrounding space. As we have seen, it is those conceptions of the world that led to the patterns of retrieval times. These conceptions of space derive from people's perceptual and behavioral experience in the spatial world. Thus, the schematic mental structures used to represent the world surrounding the body come to yield judgements that are biased. The biases are not meant to be interpreted as irrational in themselves, unless, of course, there were good general reasons why other directions should be privileged in retrieval time. Or, they may appear to be irrational if the schematization led to systematic error. The next example establishes that.

The Space of Navigation

As we have seen, the space around the body is relatively small, in reach of arms or eyes, and conceived of in three dimensions. In contrast, the space of navigation is potentially vast and is conceived of primarily in two dimensions, map-like. Unlike contemporary geographic information systems, we do not seem to carry in our minds a map of the globe that can be retrieved from particular perspectives, with requisite detail, at specific degrees of resolution. We do not seem to have such cognitive maps even for environments and maps that we know well. Rather, we seem to retrieve whatever information we can to address the problem at hand. As observed earlier, that information may be fragmented, partial, and multimodal. For actual judgements, the disparate pieces may be linked together through a schematic stucture that combines common elements, framework, or perspective.

Systematic errors give insight into the nature of the schematic structures used to link disparate spatial information. Elements or geographic entities such as buildings, roads, cities, and countries, serve as reference objects for each other. Global reference frames such as the surrounding environment or the cardinal directions also serve as referents for the location and direction of geographic entities (see Tversky, 1981; 1992).

Evidence for the use of geographic entities as reference objects comes from judgements where spatial relations to reference objects are exaggerated or simplified. For example, when people are asked to judge which map of the Americas is correct, a veridical map or one that has been altered so that South America is more directly south of North America, a significant majority pick the incorrect map. Likewise, people prefer a world map which has been altered so that the United States and Europe and South American and Africa are more aligned east and west over the real map. These errors can be accounted for by Gestalt principles of perceptual grouping. In order to perceive and comprehend scenes, people group related figures, for example in maps, geographical entities such as continents or countries or cities or roads. In memory, the spatial relations among these

figures are exaggerated or simplified, so that large comparable entities like North and South America are remembered as more aligned, in this case, north and south, than they actually are. Similar errors of alignment occur for judgements of the spatial relations among cities in these continents, for example, thinking that Philadelphia is north of Rome or Boston west of Rio. Alignment errors also occur for artificial maps and even for meaningless blobs.

Evidence for the use of global reference frames as anchors for localizing objects also comes from systematic errors of judgement. For example, to most, South America appears tilted with respect to the cardinal directions. When people are asked to place a cutout of South America as accurately as possible with respect to north-south east-west coordinates, they err in the direction of rotating South America toward upright. As for errors of alignment, errors of rotation appear in other contexts. The San Francisco Bay area is also tilted with respect to the cardinal directions and uprighted in memory so that people erroneously indicate that Berkeley is east of Stanford, and Stanford east of Santa Cruz. Similar errors occur for other actual environments, for artificial maps, and for meaningless blobs. Like alignment, this error is related to Gestalt perceptual organizing principles, in this case, common fate. The framework or natural axes of a figure are organized with respect to the framework or natural axes of the environment encompassing the figure. In memory, the natural axes of figure and ground are rotated closer in correspondence. Thus, using other entities and global axes as references for judging location and direction distorts those judgements in the direction of the reference objects and frames.

Errors of alignment and rotation are among a large number of systematic errors revealed in judgements of direction, location, and distance (see Tversky, 1992, 1993, 2000a, and 2000b for reviews). One of the more striking errors is landmark asymmetry: estimates of distances to a landmark are smaller than estimates of distances from a landmark to an ordinary building (Sadalla, Burroughs, and Staplin, 1980). This, of course, violates euclidean metrics. Together, the errors give insight into how mental representations of environments are constructed for the purposes of particular judgements. Geographic entities, such as streets, cities, or continents, or other spatial entities are organized with respect to each other and with respect to a global framework. Apparently, only the bare essentials, the schematic minima, that are needed for making the judgement are evoked. When more elements are evoked, these introduce independent constraints that serve to cancel each other and decrease errors (Baird, 1979; Baird, Merril, and Tannebaum, 1979). But when the judgements are done as they typically are, piecemeal, there is no guarantee that the bits and pieces of environments evoked to make the judgements are coherent.

These errors, alignment, rotation, and others, are a consequence of the perceptual organizing principles used to schematize the environment. If this were the only information people could use in navigation and in making spatial judgements, then people would err frequently. It is easy to conceive of situations where the errors would be consequential, for example, in deciding what direction to take to escape danger. Why is spatial information organized using principles that result in systematic error? Why didn't evolution and learning produce a system with greater accuracy and less bias?

Of course these are unanswerable questions. One can only surmise the reasons. Often,these sorts of judgements are unique to a particular time and specific place. They are

not likely to repeated, especially repeated with correction, and if they are, the corrections are made for the particular environment rather than for the general principles, the underlying system producing schematizations. Moreover, there may be natural corrections for error built into the environments. Let us now turn to route directions to see how that might happen.

Route Maps and Directions

A case that illustrates the interplay of schematic representations with corrective environments is that of route instructions. Route instructions, whether instantiated in depictions or descriptions, have a similar structure (for the structure of descriptions, see Denis, 1997; for extensions to depictions and descriptions, see Tversky and Lee, 1998, 1999). They are strings of segments containing four components: start points, reorientations, progressions, and end points. Each of these components is schematized (Tversky and Lee, 1998, 1999). This is not surprising for descriptions, as language itself is schematic. What is more surprising is that sketch maps are schematic, often even categorical, as sketch maps are potentially analog. For example, in sketch maps, paths tend to be drawn as straight or curved. The segments that are drawn as straight and those that are drawn as curved parallel the language used to direct travellers along them. For straight paths, directions use "go down." For curved paths, directions use "follow around." Intersections are typically sketched as at right angles. Where they deviate from right angles in sketches does not correspond to where they deviate from right angles in environments. Distances are notoriously incorrect in sketch maps, and vague in verbal directions. Thus, sketch maps and route directions seem to be constructed from the same underlying information, despite differences in format. Even though they are schematic, directions can be successful, even in such difficult environments as Venice (Denis, Pazzaglia, Cornoldi, and Bertolo, 1999). Why is that? Presumably because the environment itself removes the ambiguity and corrects the errors of the sketch maps and directions. For example, exact distance does not matter, as the point of change in direction, the reorientation, will be indicated by the appropriate landmark. Similarly, people may remember an intersection as at right angles, a typical consequence of alignment and rotation. If the intersection is in fact at some other angle, the traveler has no choice but to go at that angle, rather than the angle remembered. In fact, the traveler may not even notice that the turn taken is not a right-angle turn. Schematic knowledge, then, is often sufficient for successful navigation.

Why Systematic Errors Survive

The mechanisms that generate schematic representations may remain despite the fact that they can also generate systematic error. The reasons for this derive both from characteristics of the human mind and from the nature of human interaction in the world. The mind does not seem able to generate richer representations of environments both because long-term memory does not seem to contain them and because working memory is too limited to construct them. It doesn't seem to make sense for the mind to store each and every encounter with the world in great detail. Remembering the details of a particular route traveled to get somewhere doesn't help on the return home, or on another route to the same destination. Remembering the schematic structure of the route does. The same argument can be made for other kinds of information, for example, remembering faces. In fact, S. the mnemonist studied by Luria, was haunted by details of episodes, which he could not forget. As a consequence, he had difficulty recognizing patterns and generalities (Luria, 1968). For example, he was poor at recognizing faces, confused by each encounter with an individual. Thus, there are real benefits for schematic representations, not just costs. Interactions with the world do not seem to correct the mechanisms producing schematization because errors may go undetected either because of lack of feedback or because the structure of the environment corrects them. If error is detected, the correction may be specific to the context, and not to the underlying mechanisms. Readers of this paper now know that Philadelphia is south of Rome, but knowing that will probably not prevent believing that Algiers is south of Los Angeles. College students who heard an entire lecture on systematic errors in cognitive maps, including errors of alignment and rotation, persisted in making those errors even when tested immediately after the lecture.

Thus, a case can be made for rational reasons for (seemingly) irrational behavior. In some sense, it is not rational to have cognitive mechanisms guaranteed to produce error. Framed differently, however, these very same mechanisms seem reasonable, in fact, a good solution to a hard problem. Systematic errors in cognitive maps are not the only systematic errors of judgement or behavior documented in people. Nor is the analysis described here the only way to make seemingly unreasonable behavior appear reasonable. Many errors of judgement occur because the human mind is asked to measure things it has no way of measuring. Instead, it substitutes close mechanisms that it does have, for example, using availability, or the ease to which things come to mind, as a heuristic for measuring frequency (Tversky and Kahneman, 1983). Since there are other factors besides frequency that cause things to come to mind, frequency judgements based on availability will be biased by these factors.

This tension, between apparently irrational behavior on the one hand and theories of selection on the other exists not only in domains of individual behavior, where selection is by evolution as well as learning, but also in other domains where selection operates, notably survival in biological, economic, and political situations. In those arenas, too, social and biological scientists delight in discovering examples of irrationality, behaviors that are self-defeating or that their instigators would disavow. For each of these, scientists also delight in construing them as reasonable, if viewed differently. Where does this leave rationality?

References

Baddeley, A. D. (1990). *Human memory: Theory and practice.* Boston: Allyn and Bacon

Bryant, D. J. and Tversky, B. (1999). Mental representations of spatial relations from diagrams and models. *Journal of Experimental Psychology: Learning, Memory and Cognition,* 25, 137–156.

Baird, J. (1979). Studies of the cognitive representation of spatial relations: I. Overview. *Journal of Experimental Psychology: General,* 108, 90–91.

Baird, J., Merril, A., & Tannenbaum, J. (1979). Studies of the cognitive representations of spatial relations: II. A familiar environment. *Journal of Experimental Psychology: General,* 108, 92–98.

Bryant, D. J., Tversky, B., & Franklin, N. (1992). Internal and external spatial frameworks for representing described scene. *Journal of Memory and Language*, 31, 74–98.

Bryant, D. J., Tversky, B., and Lanca, M. (2001). Retrieving spatial relations from observation and memory. In E. van der Zee and U. Nikanne (Editors), *Conceptual structure and its interfaces with other modules of representation.*. Oxford: Oxford University Press.

Denis, M. (1997). The description of routes: A cognitive approach to the production of spatil discourse. *Cahiers de Psychologie Cognitive*, 16, 409–458.

Denis, M. Pazzaglia, F., Cornoldi, C., & Bertolo, L. (1999). Spatial discourse and navigation: An analysis of route directions in the city of Venice. *Applied Cognitive Psychology*, 13, 145–174.

Donald, M. (1991). Origins of the modern mind. Cambridge: Harvard University Press.

Franklin, N. and Tversky, B. (1990). Searching imagined environments. *Journal of Experimental Psychology: General,* 119, 63–76.

Luria, A. R. (1968). *The mind of the mnemonist.* N. Y.: Basic Books.

Norman, D. A. (1993). Things that make us smart. Reading, MA: Addison-Wesley.

Sadalla, E. K., Burroughs, W. J., and Staplin, L. J. (1980). Reference points in spatial cognition. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 516–528.

Tversky, A. and Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgement. *Psychological Review,* 90, 293–315.

Tversky, B. (1981). Distortions in memory for maps. *Cognitive Psychology*, 13, 407–433.

Tversky, B. (1992). Distortions in cognitive maps. *Geoforum*, 23, 131–138.

Tversky, B. (1993). Cognitive maps, cognitive collages, and spatial mental models. In A. U. Frank and I. Campari (Editors), *Spatial information theory: A theoretical basis for GIS.* 14-24. Berlin: Springer-Verlag.

Tversky, B. (2000a). Levels and structure of cognitive mapping. In R. Kitchin and S. M. Freundschuh (Editors). *Cognitive mapping: Past, present and future*. 24–43. London: Routledge.

Tversky, B. (2000b). Remembering spaces. In E. Tulving and F. I. M. Craik (Editors), *Handbook of Memory*. 363–378. New York: Oxford University Press.

Tversky, B. (In press). Spatial schemas in depictions. In M. Gattis (Editor), Spatial schemas and abstract thought. Cambridge: MIT Press.

Tversky, B., Kim, J. and Cohen, A. (1999). Mental models of spatial relations and transformations from language. In C. Habel and G. Rickheit (Editors), *Mental models in discourse processing and reasoning*. 239–258. Amsterdam: North-Holland.

Kurz, knapp, konsistent?
Schwierigkeiten mit einem regulativen Ideal

MAX URCHS

In der Tradition analytischer Philosophie ist die Forderung nach der Konsistenz von Argumentationen, Überzeugungssystemen und wissenschaftlichen Theorien ein zentrales Postulat. Die Widerspruchsfreiheit wird zum entscheidenden Kriterium für die Rationalität des Denkens und insbesondere zum Kriterium rationaler philosophischer Reflexion. Umgekehrt markieren Widersprüche das Ende jeglicher ernsthaften wissenschaftlichen Analyse.

„Widersprüche" gibt es in allen möglichen Formen und Schattierungen, der Begriff ist von schillernder Paronymität. Freilich stehen neben „Widerspruch" eine sehr große Zahl ähnlicher und oft synonym gebrauchter Begriffe zur Verfügung. Im Deutschen (ähnlich auch im Englischen und allen mir sonst zugänglichen Sprachen) gibt es, grob gezählt, 50 Varianten von Ausdrucksmöglichkeiten für alle Arten von Widersprüchlichkeit, von „absurd" über „komplementär" und „polar" bis „zirkulär". Diese können durch allerlei Kombinationen noch bedeutend vermehrt werden. Eine Klassifikation ist bislang nicht in Sicht, Regeln für einen geordneten Sprachgebrauch sind gerade erst im Entstehen. (Bislang kommt es aus rein terminologischen Gründen selbst unter Fachleuten immer wieder zu verblüffenden Mißverständnissen.) Ich werde den Begriff „Inkonsistenz" als Oberbegriff für alle Arten konfligierender oder nicht-zusammenpassender Bestandteile von Überzeugungsmengen – gleich welcher Art – nutzen. „Widersprüche" sind besondere Formen von Inkonsistenzen, bei denen eine Aussage mit ihrer (aussagenlogischen) Negation (aussagenlogisch) konjunktiv verknüpft ist. Diese Gebilde sollen, mangels einer besseren Bezeichnung, auch *et-non*-Sätze heißen. Ihre formalen Entsprechungen werden wie üblich als Kontradiktionen bezeichnet.

Da nun weder der Begriff „Widerspruch(sfreiheit)", noch „Rationalität" eine präzise Bedeutung haben, könnte man obige Thesen über Widersprüche und Rationalität einfach ignorieren – wohl jeder würde wohl zumindest einer der möglichen Lesarten zustimmen. Im Alltagsverständnis aber drücken sie sehr wohl eine klare und einflußreiche Position aus: Widerspruchsfreiheit ist der Schlußstein im Gewölbe westlicher Vernunft; fehlte die Konsistenz, so bliebe von diesem imposanten Bau nur ein Trümmerfeld übrig.

Das Bild des desaströsen Widerspruchs und die sich daraus ergebende Therapie geht auf Aristoteles zurück. Seine Auffassung erwies sich aus verschiedenen Gründen nicht nur für das wissenschaftliche Denken, sondern darüber hinaus für die abendländische Kultur der geistigen Arbeit insgesamt als prägend.

These 1: *Aristoteles prägt und paradigmatisiert das Widerspruchsverständnis der abendländischen Philosophie.*

B. Brogaard, B. Smith (Hg.), *Rationality and Irrationality / Rationalität und Irrationalität*, S. 379–389.

Schon ein kurzer Blick auf die historische Entwicklung der Widerspruchsproblematik macht deutlich, daß Aristoteles' Position zu seiner Zeit keineswegs alternativlos war. Von Heraklit zieht sich über die Eleaten eine zweite Traditionslinie, die des dialektischen Denkens, bis in die Gegenwart. Ernstzunehmende Überlegungen zum Platz und zur Rolle des Widerspruches im rationalen, insbesondere im wissenschaftlichen Denken finden sich später bei Cusanus, Kant, Fichte, Hegel, Marx ...

Eine dritte Traditionslinie, die ihren Anfang in der Sophistik nahm, ist von Anfang an explizit auch auf eine logische Behandlung des Widerspruches aus – freilich mit prinzipiell anderen Mitteln als sie Aristoteles in seiner Syllogistik angelegt hatte. Damit stand die Sophistik dem etablierten Umgang mit Mengen inkonsistenter Sätze noch ferner, als die in der Regel ohne formalen Anspruch operierende dialektische Richtung. In der unvermeidlichen Konfrontation wurde sie wirkungsgeschichtlich nahezu völlig neutralisiert. Das bedeutet freilich nicht, daß diese Ideen überhaupt aus dem geistigen Leben verschwunden wären. Die Konzeptionen der Aufklärung postulierten ein nicht zu unterbindendes Recht auf Widerspruch: nur der Abschied von der Idee der einzigen Wahrheit ebne den Weg zum gesellschaftlichen Frieden.

Insbesondere über die Literaturwissenschaft fanden derartige Positionen im vergangenen Jahrhundert auch den Weg zurück in die (kontinentale) Philosophie und machen heutzutage als postmodernistische Entwürfe anständigen, analytischen Philosophen das Leben schwer.

These 2: *Die sophistische Tradition bei der formalen Behandlung des Widerspruches verdient, neu überdacht zu werden.*

In der Tat scheint diese philosophische Option gegenwärtig eine Renaissance zu erleben. (vgl. etwa Gogos 1998)

In der Philosophie, insbesondere der akademischen, blieb die Aristotelische Ansicht zum Widerspruch paradigmatisch. Man fragt dann natürlich nach den Gründen, die zur so klaren Dominanz einer einzigen Option durch das gesamte europäische Mittelalter und die Neuzeit hindurch bis in die neueste Philosophie führten.

Aristoteles bezeichnet den Satz vom Widerspruch, wonach keinem Ding eine Eigenschaft zugleich und in der gleichen Hinsicht zukommen und nicht zukommen könne, als das sicherste Prinzip überhaupt, hinsichtlich dessen keine Täuschung möglich ist:

> Dasselbe kann demselben in derselben Hinsicht unmöglich zugleich zukommen und nicht zukommen. [Metaphysik G 3, 1005 b 19]

Seine Formulierung sichert er gegen „sophistische Belästigungen" kurzerhand ab:

> Und auch die übrigen Unterscheidungen, die als Abwehr gegen logische Schwierigkeiten hinzugefügt werden müssen, seien hinzugedacht. [ebenda, 1005]

Die logische und die ontologische Form des Satzes stellen sich, Łukasiewicz zufolge, als gleichwertig heraus (vgl. Łukasiewicz 1987, S. 15 ff.). Als logischer Satz gesehen ist nun sein (logischer) Wert an das Vorliegen eines Beweises geknüpft. Den kann Aristoteles

nicht angeben. Der Satz ist damit ohne jeden logischen Wert. Auch sein grundlegender Charakter scheint zweifelhaft. Warum ist nicht beispielsweise der doch offenbar strukturell einfachere Satz der Identität (jedes Ding ist sich selbst gleich und verschieden von allen anderen) der grundlegende, zweifelsfreie etc. Satz? Aristoteles argumentiert nicht wirklich ernsthaft für seine These, sondern schimpft vielmehr auf jene, die überhaupt entsprechende Fragen stellen. Sein Versuch, eine spezielle Beweisform, die *elenktischen Beweise*, zu nutzen, darf als gescheitert angesehen werden. Nur seine nachhaltige Autorität kann erklären, daß dieses Thema noch immer diskutiert wird. (siehe z.B. Rapp 1993, S. 521–541)

Ludwig Wittgensteins Bemerkungen über die Rolle und den Platz von Widersprüchen in den Grundlagen der Mathematik sind höchst anregend und wohlbekannt, Jan Łukasiewiczs sorgfältige Analyse des Widerspruchsprinzips ist weniger bekannt, aber kaum weniger tiefgründig und wichtig. Ohne hier Prioritäten nachzuspüren, kommt man der Wahrheit sicher nahe, wenn man annimmt, das Thema habe seinerzeit in der Luft gelegen, es war einfach reif geworden. Beide Autoren kommen zu ähnlichen Schlußfolgerungen: daß sich nämlich die Bedeutung des Satzes vom Widerspruch nicht aus logischen oder ontologischen Gründen ergibt. Die psychologische Variante zu begründen, scheint noch viel aussichtsloser zu sein. Ich halte sie einfach für falsch: es ist nicht unmöglich, miteinander Unvereinbares zu glauben. Wie sind inkonsistente Überzeugungen möglich? Nehmen wir an, einem Sprecher X könne eine Überzeugung A dann zugeschrieben werden, wenn er aufgrund seines psychischen Zustandes geneigt ist, die Frage „Glaubst Du, daß A?" zu bejahen. X hat demzufolge dann simultan die Überzeugungen A und B, wenn die latente Bereitschaft vorhanden ist, eine beliebige der Fragen „Glaubst Du, daß A? und „Glaubst Du, daß B?" zu bejahen. (Das bedeutet nicht, daß X auch bereit ist, nacheinander alle beide Fragen zu bejahen!) Man kann sich leicht vorstellen, einen Menschen zu finden, der jedem der folgenden Sätze zustimmen würde, wenn man ihn nach seiner Ansicht fragen würde: „Clinton ist ein cleverer Politiker", „Clinton hat in seiner Amtszeit als amerikanischer Präsident unverzeihliche Fehler gemacht." Diese beiden Überzeugungen sind offenbar nicht, zumindest nicht ohne weiteres, konsistent. [Über den praktischen Nachweis derartiger psychischer Zustände muß man sich als Logiker nicht den Kopf zerbrechen. Das kann man, dank Frege und Husserl, getrost den Psychologen überlassen.] Ein Anhaltspunkt, daß X in der Tat eine inkonsistente Überzeugung hat, ergibt sich aus seiner Bereitschaft, mit A und B auch $(A \rightarrow \sim B)$ bzw. $(B \rightarrow \sim A)$ zu akzeptieren. (Beide Ausdrücke sind zu $\sim(A \,.\, B)$ äquivalent.) Falls jemandem, der von obigen Sätzen überzeugt zu sein meint, der Satz „Wenn Clinton unverzeihliche Fehler unterlaufen, dann ist er eben kein cleverer Politiker" plausibel erscheint, so hat er inkonsistente Überzeugungen. Und das ist ein völlig natürlicher Zustand.

Man fragt sich unwillkürlich, ob Aristoteles vielleicht andere – als rein wissenschaftliche – Gründe hatte, mit derartiger Vehemenz für den Satz vom Widerspruch einzutreten. Solche Gründe scheint es tatsächlich gegeben zu haben. Aristoteles' Gegenspieler in der Frage des rationalen Denkens waren vor allem die Sophisten. Diese vermittelten ihr Wissen offenbar nicht immer mit dem nötigen moralischen Ernst an ihre zahlende Kundschaft. Verschiedene überlieferte Fehl- und Trugschlüsse lassen ahnen, daß es mitunter eher darum ging, den Diskursgegner gründlich zu verwirren und anschließend zu überreden, als ihn durch gute Gründe zu überzeugen.

Hier setzt nun Aristoteles an. Ihm geht es darum, in jedem einzelnen Fall der Wahrheit im Streitgespräch zum Sieg zu verhelfen. Dazu mußten Trugschlüsse als solche nachgewiesen und anschließend die sicheren Schlußregeln von den unsicheren, (be-)trügerischen unterschieden werden. Erst dann konnten in den Argumentationen auch zulässige von unzulässigen Schlußfolgerungen getrennt werden. Um diese enorme Aufgabe zu bewerkstelligen, entwickelt Aristoteles seine Syllogistik. Freilich mußte er eine Reihe vereinfachender Annahmen machen, damit das Unternehmen mit den ihm verfügbaren formalen Mitteln überhaupt beherrschbar blieb. Er mußte voraussetzen, daß bei Widerstreitendem das eine wahr und das andere falsch ist. Er mußte weiterhin annehmen, daß die Entscheidung zwischen wahr und falsch jeweils eindeutig getroffen werden kann. Er mußte dazu schließlich noch postulieren, daß es eine für jeden verbindliche und allem übergeordnete Wahrheit gibt. Erst dank dieser Annahmen konnte die Logik als eigenständige Wissenschaft entstehen. Sein Verdienst als Begründer dieser Wissenschaftsdisziplin ist unbestritten.

Aber Aristoteles zahlte einen hohen Preis: Statt der prallen, farbigen Vielschichtigkeit realer Dispute zeigte die logische Betrachtung nur noch ein erstarrtes Objekt mit idealisierten, hochgradig abstrakten Eigenschaften. Insbesondere mußte Aristoteles den Widerspruch ausblenden, um Logik in der modernen Form seiner Syllogistik überhaupt betreiben zu können.

Beim Studium philosophischer Texte vor Aristoteles findet man in der Tat eine auffallend große Zahl argumentativer Schritte, die logisch nicht korrekt sind. Man könnte also meinen, Aristoteles habe einem dringenden Erfordernis seiner Zeit entsprochen, dieser bis dato ungestraften logischen Unkultur durch Entwicklung seiner Syllogistik ein Ende zu setzen Dabei würde man freilich übersehen, daß die erwähnten Schlußfolgerungen, etwa in Platos frühen Dialogen, erst vom Standpunkt der Syllogistik aus zu Fehlschlüssen wurden! In der Geschichte der Philosophie findet man die Ansicht, die vor Aristoteles liegende Periode sei kein logikfreier Raum, sondern sie sei durch eine alternative Logik gekennzeichnet gewesen. Sicher ist eine solche These nicht unplausibel, zumindest angesichts der Alternativen: man müßte beispielsweise Plato unterstellen, er hätte schluderig gedacht, oder in seinen Dialogen allezeit nur gescherzt oder einfach Unfug geschrieben. Damit würde man aber auch Aristoteles kaum gerecht werden: die Vehemenz seines Vorgehens erschiene als völlig überzogen – die angemessene Reaktion wäre in einem solchen Fall gewesen, milde lächelnd oder schulterzuckend über die offenbar gar nicht ernst gemeinten Auslassungen hinwegzugehen. Umgekehrt wird angesichts der seinerzeitigen Lebendigkeit und Verbreitung jener Logik der absoluten Vieldeutigkeit die titanenhafte Leistung des Stagiriten erst recht deutlich.

These 1a: *Es gab keine zwingenden Gründe für die Jahrhunderte lang fast konkurrenzlose Dominanz der Aristotelischen Option.*

Aristoteles hatte sich erfolglos um einen Begründung seines „allersichersten" Prinzips bemüht. Wenn sich nun für den Satz vom Widerspruch keine logischen Gründe finden ließen, so hat er keinen logischen Wert. Das machte den Weg frei, über dessen Platz in der Logik nachzudenken. (Analoge Überlegungen lassen sich für Ontologie und Psychologie anstellen.) Zwar setzt Łukasiewicz hinzu, die Bedeutung des Satzes vom Widerspruch

entstünde aus seinem immensen ethisch- praktischen Gewicht – er folgt dabei Aristoteles, der im Satz vom Widerspruch die einzige Waffe gegen Falschheit und Lüge zu sehen meinte – aber selbst das scheint zweifelhaft. Zumindest in den Bereichen, die für den Begriff des rationalen Denkens konstitutiv sind, also in den mathematischen und empirischen Wissenschaften, haben wir durchaus andere, wissenschaftsinterne Kriterien, die Wahrheit zu bestimmen.

Man stelle sich folgende Situation vor. Zwei Experimentalphysiker, Fritz und Ferdinand, stellen bei sich divergierende Ansichten zur Siedetemperatur von Wasser fest. Fritz ist überzeugt, Wasser koche bei 80°C, Ferdinand opponiert: 90°C sei der korrekte Wert. Nun wird Fritz im Normalfall nicht zur angeblich einzigen Waffe im Kampf um die wissenschaftliche Wahrheit greifen, und dem Ferdinand den Satz vom Widerspruch um die Ohren schlagen. Sie werden vielmehr mit Thermometer, Wassertopf und Heizplatte einen kleinen Versuch durchführen, der – je nach den Luftdruckverhältnissen im Labor – einem von beiden, oder auch keinem von ihnen Recht geben wird.

Es ist allerdings entscheidend, daß Fritz auf Ferdinands Entgegnung irgendeine Reaktion zeigt, die über ein desinteressiertes Schulterzucken hinausgeht. Ein solches Desinteresse würde im Effekt Wissenschaft unmöglich machen, indem es nämlich die empirischen Disziplinen „mit Ergebnissen überlaufen ließe". Das Widerspruchsprinzip ist also durchaus doch eine Art Waffe, wenngleich von anderer Art, als Aristoteles vorgab: Eine Waffe nicht im Kampf gegen Lüge und Falschheit, sondern im Ringen um Präzision und (wo gefordert) Vollständigkeit. Wir werden beim Kommentar zur 7. These auf diesen Punkt noch zurückkommen.

Die Sache ist also offenbar – wieder einmal – wesentlich komplizierter, als dies *prima facie* der Fall zu sein schien. Wir wollen im folgenden untersuchen, wie weit man von der starren Ablehnung aller Widersprüchlichkeit in wissenschaftlichen Kontexten abrücken darf, ohne sich dem Vorwurf des Irrationalismus auszusetzen.

Die Grenzen des Zulässigen können m.E. hier durchaus weit gesteckt werden. Angesichts der Tiefe der erforderlichen Erkenntnis, in der die Inkonsistenz bestimmter Überzeugungen erst offenbar wird, ist die strikte Forderung nach Konsistenz sicherlich nicht praktikabel. Wenn sich der ontologisch dubiose Charakter des „mit Zirkel und Lineal konstruierten kreisrunden Quadrates" jedem der deutschen Sprache Kundigen erschließt, so braucht es beträchtliche mathematische Kenntnisse, um das „mit Zirkel und Lineal konstruierte Quadrat mit dem Flächeninhalt des Einheitskreises" als nahen Verwandten des ersten Objektes zu erkennen. Unbewußt widersprüchliche Überzeugungen können also offensichtlich kein Brandmal fehlender Rationalität sein, wenn wir den Rationalitätsbegriff nicht deflationär gebrauchen wollen. Aber selbst das Beharren im als solchen erkannten Widerspruch ist kein verläßliches Kennzeichen für manifeste Irrationalität. Angesichts des Umstandes, daß mitunter einfach keine konsistenten naturwissenschaftlichen Theorien vorhanden sind, können durchaus auch inkonsistente Alternativen, die aber in vielen Fällen präzise und korrekte Prognosen erbracht haben, das Mittel einer durchaus rationalen Wahl sein.

Die Grenze des rational Vertretbaren scheint vielmehr erst dort überschritten, wo man Widersprüchlichkeit als unproblematisch ansieht, sozusagen mit einem Schulterzucken darüber hinweg geht. Sicher ist auch eine solche Haltung vorstellbar. In manchen Gebieten menschlicher Geistestätigkeit scheint die gelassene Einstellung Widersprüchen ge-

genüber sogar ein grundlegendes Erfordernis für die dort noch mögliche Form von Erkenntnis zu sein. Ich denke hier vor allem an theologische Reflexionen zum Wesen Gottes, wie sie in verschiedenen Religionen anzutreffen sind. Freilich habe ich keine Schwierigkeiten mit dem Postulat, diese Reflexionen aus dem Bestand rationalen Denkens auszugrenzen. Zumindest für die christliche Religion kann man sich zu diesem Zweck auf intime Kenner der Materie stützen, die dem Verstand eben dort eine Grenze zogen, wo er sich als unfähig erwies, Gott in seiner Widersprüchlichkeit zu begreifen. Jenseits dieser Grenze ist allein die Vernunft zu weiterführender Erkenntnis fähig, diesseits der Grenze, und somit im Bereich des Rationalen, liegt etwa für Nikolaus von Kues das Reich der Wissenschaft. Dies ist das uns interessierende Gebiet rationalen Denkens. Die für das religiöse Denken typischen Widersprüche ähneln manchen künstlerischen Schwärmereien und liegen jedenfalls noch hinter dem dem Verstande zugänglichen Horizont. Sie sollen uns hier nicht weiter beschäftigen.

These 4: *Die Widerspruchsfreiheit (sogar) der exakten Wissenschaften ist ein Mythos.*

Als exakte Wissenschaften verstehe ich die Mathematik und die theoretische Physik. In der Geschichte beider Disziplinen lassen sich Phasen nachweisen, in denen die Grundlagen dieser Wissenschaften nicht konsistent dargestellt werden konnten. Für die Mathematik kann man sich natürlich auf Ludwig Wittgensteins Bemerkungen zu den Grundlagen dieser Disziplin berufen, eine geradezu verblüffend lakonische Diagnose findet sich aber auch beim jeglicher Irrationalität gewiß unverdächtigen „Nicolas Bourbaki":

> Historically speaking, it is of course quite untrue that mathematics is free from contradiction; non-contradiction appears as a goal to be achieved, not as a God-given quality that has been granted us once for all. Since the earliest times, all critical revisions of the principles of mathematics as a whole, or of any branch in it, have almost invariably followed periods of uncertainty, where contradictions did appear and had to be resolved ... There is no sharply drawn line between those contradictions which occur in the daily work of every mathematician, beginner or master of his craft, as the result of more or less easily detected mistakes, and the major paradoxes which provide food for logical thought for decades and sometimes centuries ...
> What will be the working mathematician's attitude when confronted with such dilemmas? ... Let the rules be so formulated, the definitions so laid out, that every contradiction may most easily be traced back to its cause, and the latter either removed or surrounded by warning signs as to prevent serious trouble. [Bourbaki 1949, 2–3]

Im Fall der Physik dauert ein inkonsistenter Zustand der Gesamtdisziplin weiterhin an: Die insgesamt unifizierende Theorie der Quantengravitation liegt noch immer nicht ausgearbeitet vor. Es lassen sich zahlreiche Berichte führender Vertreter der Physik über ihre Arbeit anführen, die diesen Zustand reflektieren. Bohrs Begriff der Komplementarität ist der Versuch, tradierte Vorstellungen über den unabdingbar konsistenten Charakter naturwissenschaftlicher Theorien mit den spezifischen Erfordernissen bei der Beschreibung des Untersuchungsgegenstandes in Übereinstimmung zu bringen.

Selbst in Wissenschaftsgebieten, in denen eine unübertroffen präzise Fachsprache mit höchsten Standards bezüglich methodologischer Strenge zusammenkommt, sind also Inkonsistenzen nicht immer zu vermeiden. Dann ist es freilich nicht verwunderlich, daß sich auch in anderen Gebieten der empirischen und der Geisteswissenschaften zahlreiche Beispiele inkonsistenter Theorien finden lassen.

Wenn man davon ausgeht, daß es eben die wissenschaftliche Tätigkeit der Menschen ist, die mittelfristig die Standards für rationales Verhalten etabliert, so lassen sich im Ergebnis dieser Situation, die im Selbstverständnis der betroffenen Wissenschaftler zunehmend registriert wird, für die Zukunft neue Aspekte der Rationalitätsbegriffes erwarten.

These 5: *Inkonsistenz ist effizient.*

Diese These verlangt eine Erläuterung. Nicht jede Form von Inkonsistenz ist effizient, Manche Inkonsistenzen sind ausgesprochen kontraproduktiv, irreführend, kompromittierend, etc.

Effiziente Inkonsistenz tritt nun insbesondere dort auf, wo komplementäre Beschreibungen verknappend zu einer gemeinsamen Darstellung zusammengeführt werden. Beispiele für dieses Phänomen bieten etwa Aphorismen: „Fertige Arbeit lacht!" Auch obige These ist ein Oxymoron und bestätigt sich insofern selbst. Freilich interessieren uns vor allem Inkonsistenzen in wissenschaftlichen Kontexten. Der Entstehungsmechanismus ist derselbe, wie im umgangssprachlichen Fall. „Das Licht hat Wellen- und Teilcheneigenschaften", „Furcht stimuliert und hemmt sexuelle Erregung" – in diesen Beispielen werden Aussagen, die inkompatible Gültigkeitsbedingungen haben, aus Effektivitätsgründen komprimiert und dadurch paradox.

Hier scheint auch die umgekehrte Perspektive interessant: man kann in manchen Kontexten unterstellen, daß Inkonsistenzen in der Darstellung durch solche komplementären Kompositionen entstanden sind. Dann sollte es prinzipiell auch möglich sein, die multidimensionale Form in eine Familie inkompatibler und dabei konsistenter Projektionen zu dekomponieren. Problematisch wird eine paradoxe Darstellung im wissenschaftlichen Kontext dann, wenn diese Auflösung nicht gelingt. Dies kann darin begründet liegen, daß die Aussage fehlerhaft komprimiert wurde, aus z.T. falschen Aussagen, oder daß eine Dekomposition prinzipiell nicht gelingt (weil keine konsistenten Projektionen gefunden werden), oder daß mehrere Dekompositionen möglich sind und nicht festzustellen ist, welche davon den Intentionen des Autors gerecht wird, welche also die korrekte Dekomposition ist.

In alltäglichen Situationen sind wir für gewöhnlich zu korrekten Dekompositionen durchaus in der Lage (wenngleich Irrtümer nie ausgeschlossen sind.) In wissenschaftlichen Kontexten hängt die Fähigkeit zur Dekomposition meist von der Fachkundigkeit ab. Es soll aber ausdrücklich nicht angenommen werden, die konsistente Dekomponierbarkeit sei in jedem einzelnen Fall möglich.

In der Philosophie ist die Sache ohnehin meist ungleich komplizierter. Zu viele Dimensionen lassen die Zahl möglicher Dekompositionen enorm anwachsen. Man endet also im Erfolgsfall mit einer konsistenten Dekomposition – ohne jedoch sicher zu sein, damit die ursprüngliche Situation rekonstruiert zu haben.

Aristoteles Kritik am Widerspruch richtet sich gegen die *et-non*-Widersprüche. Viele

„leichte" Inkonsistenzen hätte er sicher ganz gern durchgehen lassen. Nur entwickelt sich die durch ihn geschaffenen Logik nach eigenen Gesetzen. Angesichts einer in ihren Ausdrucksmitteln überaus beschränkten formalen Sprache kommt es zwangsläufig dazu, daß viele der lebensprallen Inkonsistenzen beim Formalisieren zu *et-non*-Sätzen plattgeklopft werden.

Eine konsistente Darstellung ermöglicht, das Dargestellte mittels herkömmlicher logischer Verfahren zu behandeln. Allerdings erfordert es mitunter beträchtlichen Aufwand, bei der Präzisierung einer Äußerung sowohl inhaltliche Adäquatheit, als auch Konsistenz zu erzielen. Eine komplizierte sprachliche Form der Darstellung stellt wiederum hohe Anforderungen an den formalen Apparat zur Modellierung der Ableitungen. Die verwendeten formalen Mittel übersteigen gewöhnlich schnell den klassischen aussagenlogischen, und selbst den Bereich der vollen Prädikatenlogik erster Stufe. Dann werden die aus der mathematischen Logik geläufigen Einschränkungen relevant, insbesondere steht kein algorithmisches Verfahren mehr zur Verfügung, die Wahrheit beliebiger formalisierter Aussagen zu verifizieren. Einfache Terminologie und Grammatik, die beispielsweise ohne temporale oder Sprecherindizes auskommen, senken also den erforderlichen modelltheoretischen und somit rechentechnischen Aufwand – sie lassen aber bisweilen keine adäquate und konsistente Darstellung zu. Da überdies die rechentechnischen Kapazitäten immer begrenzt sind, ergibt sich der aus der Künstlichen Intelligenz wohlbekannte Interessenkonflikt zwischen der Form der Darstellung und der Kompliziertheit des metamathematischen Apparates.

Die Entscheidung zwischen Komplexität und Inkonsistenz erfordert stets ein Abwägen des einzelnen Falles. Sollte man geringe Komplexität präferieren, dann möchte man die Risiken einer solchen Entscheidung abschätzen können. Angesichts der Logik-Entwicklung in den letzten fünfzig Jahren sind die Aussichten hier gut:

These 6: *Man kann Inkonsistenz kontrollieren.*

Im praktischen Umgang mit Inkonsistenzen sind wir ziemlich gut bewandert. Wir haben damit in der Regel nicht mehr Schwierigkeiten als mit anderen anspruchsvollen Inferenzsituationen, wie etwa kausalen oder temporalen Schlüssen, oder auch nur bei Ableitungen mit iterierten Negationen.

Wir haben gesehen, daß die in der analytischen Tradition geforderte Rigidität im Umgang mit Inkonsistenzen der schillernden Vielfalt, mit der Widersprüche in wissenschaftlichen Zusammenhängen auftreten, nicht gerecht wird. Insbesondere ist sie der praktisch üblichen Vorgehensweise bei der Bildung neuer Theorien nicht angemessen. Die in klassischer Reinheit verharrende Logik wird als Wissenschaftsmethodologie zunehmend weltfremd und neigt dazu, sich zu einer nur formalen, der Mathematik nahestehenden Disziplin zu verpuppen. 2.500 Jahre nach ihrer Entstehung hat aber die moderne Logik einen Stand der Entwicklung ihrer technischen Mittel erreicht, der manche der für Aristoteles noch unumgänglichen, vereinfachenden Randbedingungen verzichtbar macht.

In Weiterentwicklung des klassischen Prototyps entstand vor mehr als einhundert Jahren die moderne Logik. Insbesondere in den letzten dreißig Jahren hat diese Disziplin erstaunliche Fortschritte bei der Erweiterung ihrer formalen Ausdrucksmöglichkeiten erlebt. Die logik-kontrollierten Sprachbereiche haben sich Schritt für Schritt über die

Modallogik, Zeitlogik, die deontische, epistemische, kausale und Handlungslogik enorm erweitert. Typischerweise führte jede dieser Erweiterungen aber auch zu einem merklichen Anwachsen des technischen Apparates der entsprechenden Kalküle, der metamathematische Aufwand für die neuen Systeme nahm in einem Maße zu, welches jede praktische Handhabbarkeit in Frage stellte. Die wichtigsten Gebiete, auf denen die moderne Logik ihre Nützlichkeit nachweisen muß, sind die Wissenschaftstheorie und die Computerwissenschaften, vor allem auch die Forschungen zur künstlichen Intelligenz. Gerade hier stellt sich aber oft heraus, daß die Schwerfälligkeit der herkömmlichen Logik durch die modernen Entwicklungen noch nicht in zureichendem Maße überwunden werden konnte. Die sprichwörtliche Praxisferne des *Collegium Logicum* wird mitunter auch mit der modernen Logik assoziiert. Kurz: die Logik droht für potentielle Anwender uninteressant zu bleiben. Ein zentraler Punkt ist deshalb, daß die Logik heutzutage die Analyse ausschließlich statischer Zustände überwinden muß und insbesondere eine differenziertere Behandlung des Widerspruchs ermöglichen kann und soll.

Nicht nur beim plausiblen Schließen, sondern auch auf formaler Ebene kommen wir mit inkonsistenten Prämissenmengen inzwischen immer besser zurecht. Es genügt, sich die überaus dynamische Entwicklung der sogenannten parakonsistenten Logikkalküle anzuschauen, d.h. widerspruchstoleranter formaler Systeme.

Freilich muß man darauf achten, nicht das Kind mit dem Bade auszuschütten. Ein gelassener, nüchterner Umgang auch mit inkonsistenten Phasen bei der Entwicklung wissenschaftlicher Theorien soll nicht soweit führen, schließlich auf das Anstreben eines konsistenten Zustandes der Theorie gänzlich zu verzichten.

These 7: *Konsistenz ist ein regulatives Ideal.*

Traditionell wird – wie eingangs schon erwähnt – argumentiert, daß der Satz vom ausgeschlossenen Widerspruch der Schlußstein im Gewölbe abendländischer Rationalität sei. Ohne ihn bliebe folglich nur ein Trümmerhaufen zurück. Ein derart dramatisches Szenario ist sicher nicht plausibel. Allerdings sollte man den Satz vom ausgeschlossenen Widerspruch auch nicht ohne gute Gründe aus dem Bestand der logischen Prinzipien ausgliedern, indem man nämlich entsprechende, diesen Satz nicht enthaltende formale Kalküle wählt. Die andauernde Autorität formallogischer Grundsätze wirkt als pragmatische Leitlinie rationalen Denkens fort. Wenn man sich von bestimmten logischen Grundsätzen trennt, so kann das Auswirkungen auf die Standards dieses Denkens haben.

Überhaupt sollte man mit dergleichen Veränderungen vorsichtig sein. Mißlingende Reformen drohen stets in Beliebigkeit auszuarten. Sofern diese Beliebigkeit, etwa hinsichtlich der deutschen Rechschreibung, manchen willkommen sein mag, so wäre sie doch bezüglich eines Jahrtausende alten Konsens bei der Bestimmung rationalen Denkens weit folgenreicher.

Absolute Konsistenz ist praktisch nur selten zu erreichen. Inkonsistente Darstellungen sind unter Umständen konzentrierter und dadurch effizienter als Beschreibungen, die sich im Bemühen um weitergehende Konsistenz in Details verlieren. Andererseits gefährdet zuviel Inkonsistenz die Verständlichkeit, indem sie eindeutige Dekomponierbarkeit ausschließt. In diesem Spannungsfeld bewegt sich Wissenschaft.

Ich gebe durchaus zu, daß sich gerade in den Geisteswissenschaften mitunter Grenz-

fälle häufen. Wo aus mangelnder Achtung vor der wissenschaftlichen Wahrheit unbefangen Widersprüchliches behauptet wird, dort geht die Rationalität verloren. Dabei ist es dann ganz gleich, ob dem schlichte Schluderei zugrunde liegt, oder das überzeugte Vermeidenwollen „restriktiver Deutungshoheiten".

In Diskurssituationen, in denen inkonsistente und multipel auslegbare Ansichten vorgetragen werden, kann man, unter Verweis auf den Satz vom ausgeschlossenen Widerspruch, Präzisierungen der vorgetragenen Position einfordern. „Eben hast Du *A* gesagt, nun behauptest Du *non-A*, wie meinst Du das?" Die Rolle der Konsistenz als regulatives Ideal besteht nun gerade darin, daß über derartige Fragen nicht mit einem Schulterzucken hinweggegangen werden kann. Es mag sich dann im Folgenden immer noch herausstellen, daß auch nach bestmöglicher Präzisierung keine Konsistenz erreicht werden konnte. Aber zumindest größere Verständlichkeit wurde in der vorangehenden Debatte meist doch erzielt, die Parteien kennen ihre Standpunkte und können entsprechend der wissenschaftlichen Standards in der gemeinsamen Arbeit fortfahren. Der Satz vom ausgeschlossenen Widerspruch hat also in der Wissenschaft eine Disziplinierungs- und Ordnungsfunktion.

Wenngleich sein logischer Wert nicht vorhanden und der praktisch-ethische Wert höchst zweifelhaft ist, so besitzt das *principium contradictionis* doch erheblichen ästhetisch-kulturellen Wert. Eine nähere Analyse zeigt erfreulicherweise, daß das *ex contradictione quodlibet*-Prinzip auf Kosten eines ähnlichen, aber von ihm unterschiedenen logischen Prinzips beibehalten werden kann. Für den hinreichend flexiblen Umgang mit Inkonsistenzen genügt es, das *ex falso quodlibet*-Prinzip als nicht adäquat aufzugeben.

Die Konsistenz bleibt also regulatives Ideal rationalen, insbesondere wissenschaftlichen Denkens. Sie läßt sich nicht in jedem einzelnen Fall erreichen und nicht immer sofort. Die moderne Logik hat jedoch unterdessen die technischen Mittel bereitgestellt, die auch unter diesen Umständen ein kontrolliertes Vorgehen ermöglichen.

Die durch Aristoteles abgedrängten Traditionen können also heute in der Logik wieder aufgenommen werden. Statt postmoderner Kulturkritik oder raunendem Assoziieren „physikalischer Energieerhaltungssätze" mit symmetrischen Spiegelungen gebrochener Individuen bleibt die Logik natürlich bei ihren methodologischen Standards: es geht um präzise Begriffe und sauber explizierte Beweise, aber unter Einbeziehung neuer, dynamischer und widersprüchlicher Momente der zu analysierenden Gegebenheiten.

Moral: *Rational sein bedeutet, so konsistent wie möglich und so inkonsistent wie nötig zu sein.*

Literatur

Aristoteles, 1920 *Metaphysik*, Leipzig: Meiner.

Bourbaki, 1949 "Foundations of Mathematics for the Working Mathematician", *The Journal of Symbolic Logic* 14, 1–15 .

Gogos, G. 1998 *Aspekte einer Logik des Widerspruchs : Studien zur griechischen Sophistik und ihrer Aktualität* (Inauguraldissertation), Tübingen 173 S.

Łukasiewicz, J. 1987 *O zasadzie sprzeczności u Arystotelesa*, Warszawa: PWN.

Rapp, C. 1993 „Aristoteles über die Rechtfertigung des Satzes vom Widerspruch", *Zeitschrift für philosophische Forschung* 47, 521–541.

Rationalism and Irrationalism: The Case of Poland

JAN WOLEŃSKI

The pamphlet "Wissenschaftliche Weltauffassung. Der Wiener Kreis", the famous manifesto of the Vienna circle, published in 1929, does not mention any Polish name. The first important contact between Viennese and Polish philosophers was established when Karl Menger visited Warsaw in the autumn of 1929, although Jan Łukasiewicz met Moritz Schlick in Vienna in 1928 and discussed with him some philosophical questions, in particular the influence of mathematical logic on philosophy (see Łukasiewicz 1934. 614). Menger was very impressed by what he observed during his stay in the capital of Poland. Years later he noted (Menger 1994. 143, 145):

> As I observed during this and subsequent visits, Warsaw between the two wars had a marvellous scientific atmosphere. The interest of the mathematicians in their own as well as their colleagues' and students' work was of an intensity that I have rarely observed in other mathematical centers. I discovered the same spirit in the Warsaw School of Logic. But up to that time the Polish logicians had been somewhat isolated. [...]. So I decided to familiarize the Vienna Circle as well as the members of my Mathematical Colloquium with the logico-philosophical work of the Warsaw school and invited Tarski to deliver three lectures before the Colloquium, to two of which I planned to invite also the entire Circle.

Tarski appeared in Vienna at the beginning of 1930 and, except for the lectures mentioned above by Menger, he had discussions with Rudolf Carnap and Kurt Gödel, and he invited the former to Warsaw. Tarski's conception of metamathematics attracted Viennese logicians and Carnap delivered a special talk "Tarski und die Bedeutung der Metamathematik" in the Vienna Circle (see Bonk and Mosterin 2000. 26). Carnap went to Poland the same year and gave three lectures in Warsaw. Like Menger, he was also greatly impressed by the level of Polish philosophy. In his intellectual autobiography, Carnap wrote (Carnap 1963. 31):

> I found that the Polish philosophers had done a great deal of thoroughgoing and fruitful work in the field of logic and its applications to foundation problem, in particular the foundations of mathematics and in the theory of knowledge and the general theory of language, the results of which were almost unknown to philosophers in other countries. I left Warsaw grateful for the many stimulating suggestions and the fruitful exchange of ideas which I had enjoyed.

The contact initiated by visits of Menger and Carnap in Warsaw, and Tarski in Vienna became fairly intensive in the thirties. On the occasion of the International Philosophical Conference in Prague (1934), the Vienna Circle organized a special meeting devoted to the unity of science (Einheit der Wissenschaft. Prager Vorkonferenz der International Kongresse für Einheit der Wissenschaft) with papers by Charles W. Morris (2), Otto Neurath, Kazimierz Ajdukiewicz (2), Carnap, Hans Reichenbach, Janina Hosiasson, Ernest Nagel, Moritz Schlick, Edgar Zilsel, Philipp Frank, Tarski, Louis Rougier, Jan Łukasiewicz and Jörgen Jörgensen. The considerable number of Polish speakers (4 among 14) confirms the high opinion about philosophy in Poland in the Vienna Circle; the proceedings of the Prague meeting (published in the journal "Erkenntnis" in 1935) include special bibliographical information about works of Poles covering writings of 22 persons. Additionally, Ajdukiewicz was asked to deliver a special talk about the development of Polish philosophy related to ideas of logical empiricism; Morris did the same with respect of American philosophy. This was the genesis of Ajdukiewicz' paper "Der logistische Antiirrationalismus in Polen" (Ajdukiewicz 1935).

The title of Ajdukiewicz's paper was not accidental. It seems that he intended to stress the independence of Polish analytic philosophy (the Lvov-Warsaw school) from the Vienna Circle. Although Ajdukiewicz pointed out that Polish analytic philosophy was related and similar to logical empiricism in the basic methodological tenets (see Woleński 1989, 1989a for a more extensive treatment of the relation between both groups), he also added (Ajdukiewicz 1935. 30; page-references are to reprints if mentioned in bibliographical references at the end of this paper):

> There are in Poland no absolute adherents of the Vienna Circle. I do not know of any Polish philosopher who would have assimilated and accepted the material theses of the Vienna Circle. The affinity between some Polish philosophers and the Vienna Circle consists in the similarity of the fundamental methodological attitude and the affinity of the problems analysed.

Ajdukiewicz characterized the main tendencies of the Lvov-Warsaw School in four points: (a) antiirrationalism, that is, the postulate demanding that only such statements should be accepted which are justified by intersubjective means; in particular, antiiriationalism rejects all forms of mystical intuition or Wesenschau; (b) the postulate of conceptual clarity and linguistic exactness; (a) and (b) together establish that the value of philosophical enterprise is subjected to the same methodological criteria as those applied to special sciences; (c) Polish philosophy accommodated logical conceptual apparatus and generally became strongly influenced by formal (symbolic, mathematical) logic; (d) these general points, that is, antiirrationalism, the postulate of conceptual clarity and linguistic exactness, and the influence of logic, determined to a great extent the interests of Polish analytic philosophers who concentrated on scientific knowledge, metatheoretical and intertheoretical investigations, semantics and the foundations of deductive sciences.

Ajdukiewicz also pointed out the historical background of logical antirrationalism, mentioning few past Polish philosophers who anticipated the pattern developed by the Lvov-Warsaw school. It is an important matter, because a general and quite popular picture of Polish mentality suggests that it is far from rationalism. Thus, it is said that Po-

B. Brogaard, B. Smith (Hg.), *Rationality and Irrationality / Rationalität und Irrationalität*, S. 390–405.

land, due to the lack of suitable political reforms, lost its status as one of the leading European powers in 17th century and finally lost its independence in 18th century. Although Poles are appreciated for their boldness and military skills, they are also blamed for excessive and unrealistic political romanticism manifested, for instance, by several unsuccessful national uprisings resulting in unnecessary losses, sometimes extremely tragic, as in the case of fighting against the Germans in 1944 (the Warsaw uprising). This attitude was perhaps best expressed by Adam Mickiewicz, a Polish national poet, in his appeal to Poles: measure strength according to aims, not aims according to strength. I am very far from denying some more or less negative features of the Polish spirit. However, please note that the history of every nation is a result of internal and external circumstances. It is, for example, true that Poland rejected Hitler's claims in 1939 on the base of somewhat exaggerated imagination about national power, pride and honour, but, on the other hand, Great Britain and France, official Polish allies (by valid international treaties), promised much more that they actually did in September 1939 (the so-called paper war against Germany). We can also point out examples of perfect rationality in the behaviour of Poles in various extremely difficult moments of the national history. I will mention two such cases, both from the recent history. The first case concerns the years 1939-1944, probably the most dramatic period of the whole history of Poland. I will begin with an anecdote. In 1946, one library from Argentina asked to sent missing scientific Polish journals from the time of war. This manifests a quite general inability to understand what was going on under Nazi occupation in Poland. It is true that scientific journals were not published in Poland. However, in extreme circumstances, Poles organized clandestine universities and high schools. Not only that, but Poles created the entire underground state with its own courts and administration. Another example concerns the "Solidarity" movement which defeated communism in Poland in a fully peaceful manner. Clearly, both cases do not suggest that they were rooted in any irrationalism. On the contrary, the Polish underground state and fighting against communism should be interpreted as manifestations of rationalism of a very high rank. I will not continue this theme which appeared in my paper only incidentally in order to avoid too rash generalizations. Although this paper is not an essay in history of Poland and its national fate, if any, let me also note that philosophy very often does not display national spirit, if any. At first, I would like to show that the history of Polish philosophy does not confirm the opinion that Poles are irrational.

Philosophy appeared as an academic discipline in Poland in the 15th century; the University of Cracow became the centre of philosophical and scientific life in Poland. All important scholastic philosophical trends were represented in Cracow and logic was quite strong, although Poland at that time had no thinkers of the rank of Thomas Aquinas, Duns Scotus or Wilhelm of Ockham. Yet, one aspect of Polish thought in the late Middle Ages is worth mentioning. It is the idea of Pawel Wlodkowic (Paulus Wladimirus) that pagans cannot be converted to Christianity by force, in particular by military violence. This view considerably contributed to the notion of just war and prepared the principle of religious tolerance. Copernicus was the most important person of the Polish Renaissance, although his influence as a philosopher was limited. The most glorious ideas stemming from Poland in this period belong to political philosophy and are closely related to Reformation. Andrzej Frycz-Modrzewski proposed a general political reform of the Polish state and a new solution of the status of the Church; he was a Calvinist and demanded democracy in

religious institutions. The Socinians (Polish Brethren) were the most original product of the Reformation in Poland. Their doctrines, based on ideals of non-violence, common equality, justice and tolerance, became widely popular in Poland and outside, and influenced several thinkers, including Hugo Grotius and John Locke. Poland was also the first country in which the idea of religious tolerance found its practical realization. The second half of the 17th century was not a good time for philosophy in Poland. The Counterreformation triumphed, the Socinians had to leave the country and Catholic obscurantism became a normal standard, although some interesting logicians, like Marcin Śmiglecki (very popular in Oxford) were active. This period also brought a considerable crisis in Poland as a political organism. The Polish Enlightenment above all developed as a movement for political reforms in order to save Polish independence. Polish philosophy at that time was mainly influenced by ideas imported from France. Hugo Kołłątaj, Stanisław Staszic and Jan Śniadecki belonged to the main representatives of the Polish Enlightenment. Their ideas were closely related to French rationalism consisting in a fusion of elements of Cartesianism with British empiricism; Jędrzej Śniadecki tried to combine Kantianism with Scottish philosophy of common sense.

The first half of the 19th century is the only period in which Polish philosophy was dominated by irrationalism associated with romanticism and German-style idealism. So-called Polish national philosophy tried to establish how Poland lost its independence and how to recover it. Messianism was originated by Joseph Hoene-Wroński and developed by great Polish romantic poets (Mickiewicz, Słowacki and Krasiński) attributed to the Polish nation a special role (Poland is the Messiah of Nations), which requires its sacrifice for the salvation of humanity. This attitude, which justified national uprisings (1830–1831, 1846–1848, 1863–1864) was strongly criticized by Warsaw positivists (Julian Ochorowicz, Aleksander Świętochowski, Adam Mahrburg) very strongly influenced by August Comte, John Stuart Mill and Herbert Spencer. The members of this group recommended organic work consisting in the education of lower groups of society, rational organization of economics, etc. The enterprise of organic work was considered as a necessary condition of any successful fight for independence. These ideas became quite popular in Poland, divided at that time between Russia, Germany and Austro-Hungary.

Thus, if we look back at the development of philosophy in Poland, it is very far from the truth to say that it was dominated by irrationalism. It was quite contrary: Polish philosophy was usually sober, pluralistic and concerned with national problems. This tradition certainly influenced Kazimierz Twardowski, the founding father of the Lvov-Warsaw School. However, Twardowski's main ideas came from Brentano, his teacher. In particular, Twardowski followed Brentano in the view that "Vera philosophiae methodus alia nisi scientiae naturalis est" and made it the fundamental claim of his understanding of philosophy. The claim expressed in this statement suggested to Twardowski that philosophers should resign from studying those problems, even traditional and respectable, which could not be treated using genuine scientific methods. He sharply distinguished philosophy as science from 'world-views'; the main metaphysical positions belong, according to him, rather to the latter than to the former. Twardowski very strongly insisted upon conceptual clarity and linguistic strictness. Here we have a very characteristic declaration of this attitude to language (Twardowski 1920a. 257–258):

> The question [...] arises as to whether the lack of clarity that characterizes the style of some philosophical works is something unavoidable. [...]. [...], the basis for the opinion that it is impossible to write clearly about certain philosophical matters and issues remains a mystery. It is difficult to imagine someone being able to demonstrate that all writings which deal with particular philosophical topic are characterized by an unclear style. On the other hand, it is much easier to show that some philosophers are able to express themselves quite clearly even about subject-matter that is universally recognized as difficult and complicated. This leads to the conjecture that some philosophers' unclarity of style is not an inevitable consequence of factors inherited in the subject-matter of their expositions, but has its source in the muddled and vague character of their thinking. The situation would then appear to be as follows: clarity of thought goes hand in hand with clarity of style insofar as whoever thinks clearly also writes clearly and we would have to conclude that an author who writes unclearly does not know how to think clearly.

Twardowski also had a special vision of the development of philosophy in Poland; he understood the term "Polish national philosophy" simply as denoting the sum of results achieved by Polish philosophy. According to him, there are philosophical superpowers, like England, France and Germany, as well as philosophical provinces, where Poland belongs. It is understandable that superpowers dominate over provinces and it is very dangerous if a philosophical thought of a small nation becomes overdominated by a superpower because it loses its autonomy. Hence, philosophers working in a provincial country should try to keep a balance in their thought by bringing together ideas coming from various superpowers. Although Twardowski himself was mainly influenced by German-language philosophy, he recommended to his students British and French philosophy. Moreover, Twardowski postulated that philosophers of any nation should produce suitable textbooks in national languages. This concerns particularly the history of philosophy because historians from superpowers are inclined to exaggerate achievements of their own compatriots, neglecting ideas of other circles and usually completely ignoring ideas from provinces. Hence, if a nation wants to make its philosophy known even to its own members and show how it was and is related to other philosophies, it must be able to create a history of philosophy suitably balanced with respect to superpowers and correctly indicating products of the provinces, and particularly of ist own thought. Twardowski was a true pedagogoc genius. He trained more than 30 university professors in philosophy and other fields of humanities. Ajdukiewicz, Tadeusz Czeżowski, Tadeusz Kotarbiński, Jan Łukasiewicz, Stanisław Leśniewski and Zygmunt Zawirski belonged to Twardowski's leading students and all became representatives of logical antiirrationalism. His influence was well summarized by Alfred Tarski, a philosophical grandson of Twardowski (via Kotarbiński, Leśniewski and Łukasiewicz) (Tarski 1930. 20):

> Almost all researchers, who pursue the philosophy of exact sciences in Poland, are indirectly or directly the disciples of Twardowski, although his own work could hardly be counted within this domain.

If we compare Ajdukiewicz's characterization of logical antiirrationalism with Twardowski's metaphilosophical program we easily find common points. Admitting only those judgements that are intersubjectively controlable is equivalent to the claim that the genuine method of philosophy is the same as the method of science. Some Polish philosophers were inclined to identify the rational with the scientific. This path we find in Tarski, for example (Tarski 1954. 51):

> I used the word "rational" (as opposed to "irrational") in that wide sense in which it covers deductive and inductive methods alike.

Perhaps the most impressive statement to this effect is to be found in Ajdukiewicz's emphatic characterization of antiirrationalism (Ajdukiewicz 1949. 45–46):

> Rationalism values cognition whose paradigm is scientific cognition or more precisely whose paradigms are the mathematical and natural sciences. It rejects cognition based on revelation, all divinations, forebodings, prophecies, crystalgazing, etc. [...] Perhaps scientific cognition can be characterized best by emphasising two requirements which it must satisfy. Scientific cognition is such that only such content of thought as can be communicated to others in words is understood literally, that is without metaphors, analogies and others half-measures for the transformation of thought. Secondly, only those assertions can pretend to the title of scientific cognition whose correctness can be decided in principle by anybody who finds himself in the appropriate external conditions. In a word, scientific cognition is that which is intersubjectively communicable and controllable.

Although Ajdukiewicz did not explicitly mention the distinction between philosophy as a science and world-views, it is rather obvious that he accepted Twardowski's related view. The same concerns Twardowski's characterization of philosophy as a national enterprise. In fact, Ajdukiewicz's interests in French conventionalism arose as a result of Twardowski's insistence that Polish philosophy should not be overdominated by German language traditions.

A characteristic openness of the Lvov-Warsaw School to new ideas coming from other circles, related to Twardowski's vision of Polish national philosophy, is well evidenced by the reception of mathematical logic in Poland. Twardowski himself, as Tarski noted (see above) did not work in mathematical logic and logical analysis of the foundations of science. On the other hand, Twardowski very strongly stressed the importance of logical culture. He wrote (Twardowski 1920. 185, 193):

> I will not speak about specialized logical culture, but about the general one which every educated person should possess, just as he or she should possess a general historical, mathematical, grammatical, scientific, literary, etc. culture. General logical culture, like, for example, mathematical or logical culture consists in having some amount of knowledge and cultivation of some skills. [...]. I mentioned examples which sufficiently demonstrate a drastic lack of general logical culture in our journalism, textbooks for elementary schools, scientific literature and literary works. [...] Pointing out those lacks can be viewed as scholarly pedantry. However, I think that

> the problem concerns very important matters which have significant effects. It is because the lack of logical culture not only decreases intellectual level from the theoretical point of view, but also brings ignorance and obscurity into the practical applications of our thoughts. Needless to say that the whole of our life is that practical application.

Twardowski's metaphilosophical project and his insistence on logical culture contributed to the fact that his students were vitally interested in logic in the wide understanding, covering semiotics, formal logic and methodology of science. This was by no means a peculiarity of Polish philosophy, because several other circles were equally sensitive to general logic. The rise and development of mathematical logic in Poland deserves special attention. Since I tried to analyze this phenomenon elsewhere (see Woleński 1989, 1995), here I will restrict myself to very general remarks about the rise of the Warsaw group of logicians with merely incidental remarks about other places, of which Cracow was a quite strong centre of logic (Leon Chwistek, later in Lvov and Jan Śleszyński). In the academic year 1899–1900 Twardowski delivered in Lvov a course about new directions in logic. The course was mainly devoted to Brentano's reform of traditional logic, but also informed about rudiments of the algebra of logic. This course was attended by Łukasiewicz, who became interested in the new logic and began to lecture in Lvov in 1906. In 1910 Leśniewski joined Twardowski's circle. After the reopening Warsaw University in 1915 Łukasiewicz was appointed professor of philosophy (from 1920 his position was at the Faculty of Mathematics and Science), but was mainly teaching logic. In 1919 Leśniewski was given the chair of philosophy of mathematics in Warsaw. The main impetus for the development of logic in Warsaw came from mathematicians. The years 1914–1918 were the period of looking for ways of possible scientific research by various academic communities of Polish scholars. Mathematicians decided (the Janiszewski program) to concentrate on set theory and topology and their applications to other fields of mathematics. This project located mathematical logic and the foundations of mathematics in the very centre of mathematics. The appointment of Leśniewski and Łukasiewicz as professors was a conscious decision of mathematicians who were not afraid to have in their company two logicians with a philosophical background. Łukasiewicz had pedagogical and organizing abilities of Twardowski's rank and very soon succeeded (together with Leśniewski) in creating a group of mathematical logicians in Warsaw. Tarski, one of the greatest logicians in the history of logic became the third pillar of this group. The first ten years (1920–1930) constituted a period of building up the Warsaw school of logic and achieving the first important results. The school became internationally famous in the thirties; the above mentioned contacts with Viennese logicians, mathematicians and philosophers were significant for introducing Poles into the international logical forum. Thus, Polish logic (to use this not very happy label because logic is neither Polish, nor German, etc.) started almost from nothing at the the end of 19th century and during the life of one generation (thirty years from 1900–1930) achieved the top of the world in this field. Probably nobody could have predicted this development.

Several factors contributed to this glorious career of logic in Poland. Twardowski's tradition was one of them. It is not known whether Zygmunt Janiszewski who elaborated the program of the development of Polish mathematics was influenced by Twardowski's

vision of national philosophy, but it is clear that Janiszewski's project was in the spirit of Twardowski, because it recommended to take what was new and promising from the international labour of scientific research. The protection of logic by mathematicians was the next important circumstance (see also the further remark below). Many young mathematicians, notably Tarski, felt free to choose logic as their main field; in fact, relations between logicians and mathematicians became worse in the thirties, but the Warsaw school of logic was too powerful at that time and did not need special protection from mathematicians. The mentioned skills of Łukasiewicz, and Leśniewski's fame as one of the most original minds in Poland also made themselves felt. The students of mathematics at Warsaw University could attend several courses in mathematical logic from elementary to very advanced. The students had to take one elementary course, but the rest was optional. Logic teaching was organized in a way that students who wanted to specialize in logic or at least to widen their logical knowledge, could attend logical courses and seminars during the whole period of studies, that is, four years. The fact that "Principia Mathematica" served as a textbook for advanced students of mathematics gives an impression of how logic was taught in Warsaw (by the twenties the "Principia" was regarded as an obsolete work); the teaching of mathematical logic in Poland outside Warsaw was perhaps not so intensive, but still obligatory at other Polish universities. The Warsaw group of logicians was the largest circle of people working together in a single place (eleven persons in the thirties). It is also interesting to note that Poland in the interwar period had four positions in mathematical logic while the rest of the world had only one (Münster in Germany). The Warsaw school of logic had its own scientific ideology. It was expressed by Łukasiewicz in the following way (Łukasiewicz 1929. 20, 21–22):

> Mathematical logic in Poland, particularly in Warsaw is today a very vital cell of Polish creative scientific work [compare Menger's opinion quoted above – J. W.]. [...] By a happy coincidence, philosophers and mathematicians co-operated in forming Polish mathematical logic. This fact favourably augurs the future development of this field in Poland. Mathematicians will not allow logic to be changed into a philosophical speculation, but philosophers will defend it against a slavery application of mathematical logic resulting with its restriction to an auxiliary mathematical discipline.
>
> As a matter of fact, mathematical logic is considered in Warsaw as an autonomous science having own aims and problems. [...] In Germany things look differently. I have an impression that mathematicians of Hilbert's school working in logic treat it just as an auxiliary, but not independent science. [...] On the other hand, there is no danger, I think, that Polish mathematical logic will wend anytime wrong into philosophical speculation. Polish mathematicians, who co-operate with us, think too soberly to be subjected to unscientific phantasies. This danger is much more actual in the case of German mathematicians. [...] Both mathematicians and those philosophers who began to work in mathematical logic brought with them a mature feeling of scientific precision. Almost all philosophers working in mathematical logic in Poland are students of Prof. Twardowski and belong to so called "Lvov philosophical school" where they learned how to think, clearly, responsibly and methodically. Due to this fact, Polish mathematical logic achieved much higher level of scientific precision.

The main points of Łukasiewicz's characterization of mathematical logic in Poland are these. Firstly, Polish mathematical logic is a product of co-operation of mathematicians and philosophers. Secondly, it is autonomous and despite using mathematical methods it is independent of mathematics; its coming into being is associated mainly with its philosophical background. The claim that logic is independent of mathematics as well as of philosophy seems exaggerated today, but it was very productive in the thirties in Poland. Perhaps the most important fact in this respect was that most mathematicians in Warsaw shared this view about the autonomy of logic. Certainly, it was a peculiarity of the atmosphere concerning logic in Warsaw which became crucial for the growth of a powerful logical school. For instance, the situation in Cracow was completely different, because Cracow mathematicians regarded mathematical logic as something located on the margins of mathematics, not in its centre. It prevented the development of the school of mathematical logic in Cracow, although still in 1918 this city had more advanced mathematical logicians than Warsaw. Thirdly, mathematics protects logic against speculation, but philosophy protects it against a pure "mathematization". Fourthly, mathematical logic in Poland achieved an incomparable level of precision. The last point is important for the philosophical significance of logic. According to Łukasiewicz, the level of precision required and pursued in Polish mathematical logic should serve as a paradigmatic pattern for philosophers. In this way mathematical logic contributed to that factor of logical antiirrationalism which consisted in conceptual clarity and linguistic factors. Of course, logic also contributed to Polish philosophy in another way. Many-valued logic, Leśniewski's systems and the semantic conception of truth, perhaps the most glorious results of Polish logic, clearly have a double logico-philosophical character, in particular, an obvious philosophical motivation and far-reaching philosophical consequences. Also such philosophical constructions as Ajdukiewicz's radical conventionalism, his semantic epistemology and Kotarbiński's reism were invented under a very strong influence of logical ideas. Polish philosophers, at least those logically oriented, considered logic as a source of insights collectively forming logical antiirrationalism. Perhaps here is a proper place to indicate that Polish antiirrationalism was much wider in the interwar period than only its logical manifestation. Many (even most) of Twardowski's students did not work in mathematical logic, but all accepted his antiirrationalistic metaphilosophy. A particularly impressive sign of antiirrationalism (in this case influenced by Polish logic) was the rise of the Cracow Circle, a group of Catholic philosophers (Józef M. Bocheński was among them) who were ready to introduce considerable modifications into Thomism and theology, in order to keep rational the standards propagated by Twardowski's school. Chwistek, a logician not belonging to the Lvov-Warsaw School, was another example of a antiirrationalistic philosopher, influenced by logic and social ideals.

Like Twardowski, Polish logicians were entirely convinced about the social importance of logic. Tarski touched this question in his popular textbook of logic (Tarski 1941. XIII–XIV):

> I shall be very happy if this book contributes to a wider diffusion of logical knowledge. [...] For logic, by perfecting and by sharpening the tools of thought, makes men and women more critical – and thus makes less likely their being misled by all the pseudo-reasonings to which they are incessantly exposed in various parts of the world today.

For Tarski (Henry Hiż's personal communication), "Religion divides people, logic brings them together". Łukasiewicz considered logic as the morals of speech and thought. Since insistence on logical culture was a typical feature of Polish antiirrationalism, I will quote other similar declarations (respectively, Czeżowski 1969a.190, Kotarbiński 1951. 544, Ajdukiewicz 1959. 322):

> The most important point is this. Logical culture increases the level of demands, required by society, with respect to clarity and the proper justification of statements. Due to that, declarations appealing only or mostly to emotional reactions lose their hearing in people. Journalism, political games and every public activity, if it is undertaken in order to influence society, has to satisfy these higher demands of logical correctness. It favours the seriousness of discussions, demagogues lose their footing, agreements become easier because the basis for solving controversies is found in logic. Societies having high logical culture are more undivided and compact, due not to coercion but to beacons coming from logic and protecting people against entering in erroneous paths of passions and destruction.
>
> People who are not competent in logic are inclined to chaotic thought and speech. They reason about many things at once, do not preserve a planned succession of wrongly articulated problems, and are not able to distinguish close but different matters.
>
> The ability of logically correct thinking protects not only against errors with their all painful practical consequences, but also against being suggested by slogans with the empty content, although full of emotional furniture, that are neither true nor false. Due to their emotional character, they can execute prevailing but uncontrolled by deliberation influence on human conduct.

And one more quotation which can be regarded as a generalization of expectations concerning logic in relation to antiirrationalism in general (Ajdukiewicz 1949.49):

> [...], the voice of the rationalist [that is, antiirrationalist – J. W.] is a sound social reaction, it is an act of self-defence of society against the dangers of being dominated by uncontrollable forces among which may be both a saint proclaiming a revelation as well as a madman affirming the products of his sick imagination and finally a fraud who wants to convert others to his views for the sake of his egoistic and unworthy purposes. It is better to rely on the safe but modest nourishment of reason than, in fear of missing the voice of 'Truth', to let oneself be fed with all sorts of uncontrollable nourishment which may more often be poisonous than healthy and beneficial.

These quotations explain why the teaching of logic was seen as so very important in Poland; of course, general antiirrationalism in Ajdukiewicz's sense was closely related to logic. I already mentioned how teaching of logic was organized for students of mathematics. Principally, the same concerned students of philosophy in Warsaw. In fact, logical courses and seminars conducted at mathematical studies were also directed to students of philosophy; in other Polish universities, mathematical (or formal) logic was at least obligatory for philosophers. However, teaching of logic in Poland was not restricted only

to mathematicians and philosophers. Logic was taught in all secondary schools as a part of the propedeutics of philosophy and covered at least one semester. Further, logical courses were given at all pedagogical colleges. Also all students at universities having so-called "Main Problems of Philosophy" courses took logic as a central part of this course. The level of teaching was very high. Let me briefly describe the logical content of Ajdukiewicz's textbook (Ajdukiewicz 1938): I. On concepts and propositions; 1. Psychological meaning and linguistic meaning; 2. Sentence and proposition; 3. Name and concept; 4. The scope of name and concept; 5. The content of name and content; 6. Relations between scopes of concepts; 7. Logical division; 8. Inaccuracies of speech; 9. Definition; II. Information about formal logic; 10. The principle of contradiction and the principle of excluded middle; 11. Conditional sentence and the relation of entailment; 12. The logical square and principles of conversion; 13. About classical syllogism; III. Kinds of inference. Argumentation; 14. Deductive inference; 16. Reductive inference; 17. Inductive inference; 18. Argumentation and its kinds; 19. Errors in argumentation; IV. On science; 19. The classification of science; 20. A priori sciences; 21. Natural sciences; 22. Historical science. 23. The value of science. The whole book has 214 pages and consists of three parts: A. Psychology of cognitive processes (52 pages); B. Logic (102 pages); C. Human conduct (60 pages). The proportions of particular parts clearly show how logic prevailed in the program of teaching philosophy in secondary schools. Tarski 1941 was based on a booklet published in Polish in the thirties and served as the textbook in pedagogical colleges; it covers propositional calculus, predicate calculus and elements of metamathematics. Kotarbiński's "Elements" (Kotarbiński 1929) was the textbook for a general course in "Main Problems of Philosophy"; it covers semantics, propositional calculus, the Leśniewski ontology and a lot of philosophy of science.

The power of Polish logic and extensive teaching of logic at all levels of education (except elementary) resulted with a great prestige of this field in Poland. We have several signs of that. One of them is the mentioned number of positions in logic at Polish universities. It was almost unconceivable that logic would not be taught in secondary schools and universities. It was taught during the war in clandestine education system. Polish military forces fighting in North Africa, Italy and other Western fronts organized special schools for children of officiers and soldiers. We have an almost incredible document, namely the rotaprinted edition of Ajdukiewicz 1938 (a textbook for secondary schools) published by the Publishing Section of the Second Polish Corp (a unit of Polish troops fighting in Western Europe); it was printed in Bari in 1945. Łukasiewicz spent the last months of the war in Westphalia. Just after this region was taken by the Allies, he was moved to a camp for Poles liberated from Germany and demobilized soldiers. A Polish secondary school was immediately organized in this place and Łukasiewicz taught logic there. The teaching of logic was continued in Poland after 1945. After a few years, the propedeutics of philosophy was cancelled (it did not conform with the Communist plan of education), but logic itself was preserved. After some interruption in the early fifties, it was reintroduced to universities as an optional subject. It happened that over ninety percent of students chose courses of logic. And the last illustrative fact. At the peak of the Stalin era, Ajdukiewicz was able to convince Marxists that Poland, due to its tradition in logic, should have its own professional logical journal; the first volume of "Studia Logica" appeared in 1953.

The above mentioned facts clearly show that Polish logicians succeeded in making logic a prestigious and popular subject in Poland. Several generations of Polish intelligentsia were subjected to an intensive logical training. However, it is a separate problem whether it fulfilled the expectations of Twardowski and his followers concerning an inprovement of logical culture and social effects thereof. To answer this question is difficult, if possible at all, because human conduct is governed by many factors and nobody knows how much logic participates in causing human actions. Even if we will qualify views of the Lvov-Warsaw School concerning the social role of logic and its teaching as too naive (personally, I think that they were such, or at least exaggerated), it is perhaps interesting to try to give an answer. It is relatively simple to assess the role of logical culture in particular fields, especially mathematics and philosophy. I will leave mathematics aside and concentrate on philosophy. It is doubtless that logic influenced Polish philosophy very much. The case of the Lvov-Warsaw School itself is obvious – it produced a sort of logical philosophy; the same concerns Leon Chwistek. I earlier mentioned an interesting case of Catholic philosophy in Poland. The influence of logic on Polish Catholic philosophy was not restricted to the Cracow Circle and is still alive today. Roman Ingarden was a great enemy and critic of the Lvov-Warsaw School and of applying formal logical methods in philosophy. However, his writings are much clearer than many other phenomenologists and I think that it was at least partly created because he lived in a logical environment and understood that he should respect defined methodological standards. Perhaps Polish postwar philosophy is a sort of experimentum crucis here. Simply speaking, the level of Polish academic philosophy, due to its professional, above all, logical culture was sufficiently strong in order to defend it against corruption by Marxism (see Woleński 1992). Even more, Polish Marxism adopted professional standards and became a normal philosophy (in most cases) very long before the end of communism as a political system in Poland in 1989. The fact that postmodernism and similar currents are not very popular in Poland, at least among philosophers, is due to the strength of logical culture. Briefly, logic essentially contributed to logical antiirrationalism, the latter to general antiirrationalism, and the last essentially determined the general climate of Polish philosophy. Although I am personally pleased by this fact, I must add I fully recognize that not everybody needs to be happy with this situation. And if I say that postmodernism is not popular in Poland, I do not mean that it is completely absent in Polish philosophical life. The Polish philosophical scene was always pluralistic and antiirrationalism did not change this fact. This is probably a further rational feature.

It is much more difficult to measure effects of antiirrationalistic culture in such social areas as political life, journalism, literature, etc. However, I think that we can point out some possible effects of a relatively high logical culture of the Polish intelligentsia, particularly in the prewar period. I mean the quality of Polish legal statutes or the language of humanities. It is also possible that a very rational organization of the Polish underground state in 1939–1945 was caused (partly, of course) by antiirrationalism, because people involved in those activities were educated in the antiirrationalistic tradition. Another example is "Solidarity". Although workers initiated this movement, its organizational framework, in particular, the principle of self-limiting revolution, was elaborated by intellectuals, also sensitive to antiirrationalism. These explanations are very very tempting, but I suggest them with a considerable caution and many reservations (I will

not enter into details). It is also possible that we can measure some results of decreasing logical education. Poland is also an example in this respect, because logical and philosophical teaching became limited in the last years; in secondary schools, it is absent (not only recently, but since the sixties) and much reduced in the universities. Accordingly, we observe a radical decrease of the quality of statutes, public debates and an increase of irrational, for example, religious motivations. Once more, I do not insist that these facts are only caused by limitations of logical and philosophical teaching, but this factor should be also taken into account.

The case of antiirrationalism is also interesting for purely philosophical issues, especially for the concept of rationalism. We have two oppositions: (a) rationalism versus empiricism, and (b) rationalism versus irrationalism. The immediate question is whether the word "rationalism" has the same meaning in (a) and (b). Clearly not because irrationalism is different from empiricism. On the other hand, rationalism as antiirrationalism appeared in the Enlightenment and consisted in a fusion of Cartesian themes (knowledge should be clear and well justified) with empiricism claiming that knowledge should be based on experience. It could be taken as evidence that rationalism of the French philosophers of the 18th century was a successful overcoming of a sharp contrast between rational knowledge and empirical knowledge, opposition to which goes back to ancient philosophy with its distinction between certain episteme and probable doxa; the Enlightened reason inherits in this picture virtues attributed to knowledge by rationalists as well as empiricists. However, the matter is much more complicated and requires a closer analysis of irrationalism and its place in the history of philosophy. This was done in Poland by Izydora Dąmbska (see Dąmbska 1937) and Tadeusz Czeżowski (see Czeżowski 1969b). Dąmbska distinguished the following kinds of irrationalism in philosophy: methodological (Dąmbska used the label "logical" in this case, but I changed it to "methodological", in order to avoid a confusion with logical antiirrationalism in the Ajdukiewicz understanding), epistemological, metaphysical and psychological. Logical irrationalism is a property of sentences. A sentence is irrational if and only if it is contradictory or essentially undecidable (this last property is related to the views of the Vienna Circle). However, Dąmbska adds that sentences are not irrational per se, but only if they are used with assertion. Epistemological irrationalism is a view, which legitimizes irrational sentences by pointing out various cognitive activities that are (a) other than rational ones (Dąmbska understood rationalism as Ajdukiewicz did, that is, as antiirrationalism) ones, and (b) infallible. Metaphysical irrationalism (Heraclitus, Bergson) regards reality as irrational, that is, as having a residuum, which cannot be captured by conceptual resources. Finally, a person who is inclined to use of irrational devices in justification of his or her judgments is a psychological irrationalist. Dąmbska's main concern was the relation of logical irrationalism and epistemological irrationalism to scientific knowledge. According to her, both logical irrationalism and epistemological irrationalism are inconsistent with science. This verdict concerns "questio ruris". It sometimes happens that science is ("quaestio facti") irrational but (usually) only temporary, because irrational elements are eliminated by further scientific discoveries. Czeżowski characterized traditional (he said "ancient") rationalism by the following points: (1) genuine knowledge is certain and necessary; (2) certain knowledge (episteme) directly refers to the world, more precisely to its general and necessary features; (3) apodictic evidence is the criterion of knowledge; (4) deduction is the only method of constructing science. I think that this picture must be supplemented in (2) and (4). It is certainly not enough to say that, according to rationalism, episteme directly refers to the world. We should add that episteme always was conceived as having its own special objects, Platonic Forms, Aristotelian essences, the God of the schoolmen, Husserlian eidos, etc. It is also not enough to say that deduction is the only method of constructing science. We should add that, according to rationalism, the system of knowledge constitutes deductive-assertoric theory (see Ajdukiewicz 1960), that is, theory which is based on axioms unconditionally asserted. Historically speaking, this type of rationalism very far exceeds antiquity and can be found in every period of the development of philosophy. Modern empiricism (the qualification "modern" is important, because this form of empiricism did not appear before Hume, at least in its mature form) rejects all points (1)–(4). It is not surprising and the characterization of empirical knowledge by means of negations of (1), (2), (3) and (4) is now almost standard. A further analysis of rationalism leads to perhaps unexpected conclusions. Let us start with (2) as it was supplemented. If the claim that episteme has its own object, different from empirical phenomena is taken seriously, a special faculty has to be defined in order to directly catch Forms, essences, etc. Rationalists since Plato to Husserl proposed various solutions under the same heading: intuition. Usually rational intuition was contrasted with mystical faculties. Yet rationalists always had difficulties with intersubjectivity of their intuitive knowledge (compare Husserl's dramatic attempts to prove that phenomenological intuition is intersubjective). This suggests that classical rationalism is irrationalism in spite of declarations of its leading proponents; it is epistemological irrationalism in Dąmbska's sense. A remarkable fact is that Ajdukiewicz excluded *Wesensschau* (certainly having in mind Husserlian intuition) from the scope of rational knowledge. By the way, it is now clear why Husserl so strongly insisted that the traditional concept of experience is too narrow. Ajdukiewicz himself (see Ajdukiewicz 1949. 48–49) limited irrationalism only to its classical forms, that is, mainly mysticism. However, a closer analysis of classical rationalism from the point of view of the definition of rational knowledge as intersubjective, inevitably leads to the conclusion that it was a form of irrationalism. The word "rationalism" is ambiguous and rationalism as opposed to empiricism should be differently labelled, for instance, by the word "apriorism", used in fact by Ajdukiewicz. Thus, it is not surprising that typical apriorism falls under epistemological irrationalism, because it proposes special kinds of knowledge that lack intersubjective character in all historically available descriptions of aprioristic epistemology. However, a surprising thing is that some philosophers, like Plato or Hegel, regarded reality as perfectly ordered by a special logic of being (metaphysical rationalism) and yet they proposed epistemological irrationalism. I claim that this curiosity is always present when logical analysis in the proper sense is neglected. Polish logical antiirrationalism can be regarded as an attempt to tame epistemological irrationalism by pointing out why devices proposed by irrationalism devastate reliable knowledge.

References

Ajdukiewicz, K. 1935 "Der logistischer Antiirationalismus in Polen", *Erkenntnis*, V, 151-161; repr. in D. Pearce and J. Woleński (eds.), *Logische Rationalismus. Philosophische Schriften der Lemberg-Warschauer Schule*, Franfurt am Main: Athenäum 1988, 30–37.

Ajdukiewicz, K. 1938, *Propedeutyka filozofii* (Propedeutics of Philosophy), Lwów: Książnica-Atlas.

Ajdukiewicz, K. 1949 *Zagadnienia i kierunki filozofii* (Problems and Theories of Philosophy), Warszawa: Czytelnik; Eng. tr. by H. Skolimowski and A. Quinton, Cambridge: Cambridge University Press 1973.

Ajdukiewicz, K. 1959 "Co może zrobić szkoła dla podniesienia kultury logicznej uczniów?" (What Can the School Do for Increasing the Logical Culture of Students?), *Nowa Szkoła* 2, 2-8; repr. in K. Ajdukiewicz, *Język i poznanie* (Language and Knowledge), v. 2, Warszawa: Państwowe Wydawnictwo Naukowe 1964, 322–331.

Ajdukiewicz, K. 1960 "Axiomatic Systems from the Methodological Point of View", *Studia Logica*, 205–218; repr. in K. Ajdukiewicz, *The Scientific World-Perspective and Other Essays 1931–1963*, D. Reidel, Dordrecht 1978, 282–294.

Bonk, Th. and Mosterin, J. 2000 "Einleitung", in R. Carnap, *Untersuchungen zur allgemeinen Axiomatik*, Darmstadt: Wissenschatfliche Buchgesellschaft, 1–52.

Carnap, R. 1963 "Intellectual Autobiography", in P. Schilpp, *The Philosophy of Rudolf Carnap*, La Salle: Open Court, 3–84.

Czeżowski, T. 1969 "Kilka uwag o racjonalizmie i empiryzmie" (Remarks on Rationalism and Empiricsm), in T. Czeżowski, *Odczyty filozoficzne*, Toruń: Towarzystwo Naukowe w Toruniu, 18–22.

Czeżowski, T. 1969a "O kulturze logicznej" (On Logical Culture), in T. Czeżowski, *Odczyty filozoficzne* (Philosophical Essays), Toruń: Towarzystwo Naukowe w Toruniu, 185–190.

Dąmbska, I. 1937 "Irracjonalizm a poznanie naukowe" (Irrrationalism and Scientific Knnwledge), *Kwartalnik Filozoficzny* 14, 83–118, 185–212.

Kotarbiński, T. 1929, *Elementy teorii poznania, logiki formalnej i metodologii nauk*, Lwów: Ossolineum; Eng. tr. (as *Gnosiology. The Scientific Approach to the Theory of Knowledge*) by O. Wojasiewicz, Oxford: Pergamon Press 1966.

Kotarbiński, T. 1951 "Zadania swoiste logiki szkolnej" (Special Tasks of Logic in Schools), *Nowa Szkoła* 5, 327–339; repr. in T. Kotarbiński, Wybór pism (*Selected Writings*, v. 2, Warszawa: Państwowe Wydawnictwo Naukowe, Warszawa 1958, 533–552.

Łukasiewicz, J. 1929, "O znaczeniu i potrzebach logiki matematycznej" (On the Significance and Needs of Mathematical Logic), *Nauka Polska* X, 604–620.

Menger, K. 1994 *Reminiscences of the Vienna Circle and the Mathematical Colloquium*, Dordrecht: Kluwer Academic Publishers.

Tarski, A. 1930 "Brief an Otto Neurath" (25.IV.30), *Grazer Philosophische Studien* 43, 1992, 10–12; Eng. tr. by J. Tarski, *Grazer Philosophische Studien* 43, 1992, 20–22.

Tarski, A. 1941, *Introduction to Logic and to the Methodology of Deductive Sciences*, Oxford: Oxford University Press.

Tarski, A. 1954 "[Contributions to the Discussion]", *Revue Internationale de Philosophie* 8, 51; repr. in A. Tarski, *Collected Papers*, vol. IV: 1958–1979, Basel: Birkhäuser 1986, 716.

Twardowski, K. 1919-1920, "O wykształcenie logiczne" (Towards Logical Education), *Ruch Filozoficzny* V, 65-71; repr. in K. Twardowski, *Rozprawy i artykuły filozoficzne* (Philosophical Papers and Essays), Lwów: Nakładem Uczniów 1927, 185–193.

Twardowski, K. 1920, "O jasnym i niejasnym stylu filozoficznym" (On Clear and Unclear Philosohical Style), *Ruch Filozoficzny* V, 23–25; Eng. tr. by A. Szylewicz, in K. Twardowski, *On Actions, Products and Other Topics in Philosophy*, Amsterdam: Rodopi 1999, 257–259.

Woleński, J. 1989 *Logic and Philosophy in the Lvov-Warsaw School*, Dodrecht: Kluwer Academic Publishers.

Woleński, J. 1992, "Philosophy inside Communism: the Case of Poland", *Studies in Soviet Thought* 43, 93–100.

Woleński, J. 1995 "Mathematical Logic in Poland 1900-1939: People, Institutions, Ideas", *Modern Logic* 5, 363–405; repr. in J. Woleński, *Essays in the History of Logic and Logical Philosophy*, Jagiellonian University Press, Kraków 1999, 59–84.

Autorenverzeichnis
List of Authors

Prof. Dr. David M. **Armstrong**
Sydney University
Philosophy Department
New South Wales (NSW), Australia 2006
david.armstrong@philosophy.usyd.edu.au

Dr. Michael **Beaney**
Open University
Department of Philosophy
Milton Keynes MK7 6AA
M.A.Beaney@open.ac.uk

Prof. Dr. Anat **Biletzki**
Tel Aviv University, Lester and Sally Entin Faculty of Humanities
Department of Philosophy
P.O.B. 39040, Ramat Aviv, 69978 Tel Aviv, Israel
anatbi@post.tau.ac.il

Prof. Dr. Berit **Brogaard**
Rochester Institute of Technology
Department of Philosophy
Rochester, New York
brogaar@attglobal.net

Prof. Dr. Andrzej **Bronk**
Catholic University of Lublin
Department of Philosophy of Science
ul. Jagiellońska 45, PL 20-950 Lublin, Poland
bronk@kul.lublin.pl

Prof. Dr. Stephen **Clark**
University of Liverpool
Department of Philosophy
Liverpool L69 3BX
srlclark@liverpool.ac.uk

Prof. Dr. Ronald **De Sousa**
University of Toronto
Department of Philosophy
215 Huron Street, Toronto / Ontario / Canada
sousa@chass.utoronto.ca

Prof. Dr. Luis **Flores H.**
Pontifica Universidad Católica de Chile
Instituto de Filosoffa
Av. Jaime Guzmán 3300, C.P. 6650008
Santiago, Chile
lfloresh@puc.cl

Prof. Dr. Lynd **Forguson**
University of Toronto
Department of Philosophy
15 King's College Circle, Toronto, Canada M5S 3H7
lynd.forguson@utoronto.ca

Andrew U. **Frank**
Geoinformation und Landvermessung
Technische Universität Wien
Gusshausstr. 27–29 / 127
frank@geoinfo.tuwien.ac.at

Prof. Dr. Newton **Garver**
University at Buffalo
Department of Philosophy
Buffalo NY 14260-4150
garver@acsu.buffalo.edu

Dr. Ivan M. **Havel**
Center for Theoretical Study
The Institute for Advanced Studies at Charles University and
the Academy of Sciences of the Czech Republik
Jilská 1, 110 00 Praha 1, Czech Rep.
havel@mbox.cesnet.cz

Prof. Dr.Herbert **Hochberg**
The University of Texas at Austin
Department of Philosophy
Waggener Hall 316, Austin, Texas 78712-1180
hochberg@mail.utexas.edu

Prof. Dr. Terry **Horgan**
University of Memphis
Department of Philosophy
Memphis, TN 38152

Prof. Dr. John **Kearns**
University at Buffalo, the State University of New York
Department of Philosophy and Center for Cognitive Science
131 Park Hall, Buffalo, New York 14260, USA
kearns@acsu.buffalo.edu

Prof. Dr. Edgar **Morscher**
Universität Salzburg
Institut für Philosophie
Franziskanergasse 1, A-5020 Salzburg
Edgar.Morscher@sbg.ac.at

Prof. Dr. Philippe **Nemo**
ESC-EAP
Graduate School of Management
79 avenue de la République, 75543 Paris
phnemo@worldnet.fr

Prof. Dr. J. C. **Nyíri**
University of Budapest (ELTE)
Institute for Philosophical Research, Hungarian Academy of Sciences
H-1054 Budapest, Szemer u. 10
nyiri@phil-inst.hu

Prof. Dr. mr. Herman **Philipse**
Faculty of Philosophy
University of Leiden
Postbus 9515, 2300 RA Leiden, The Netherlands
philipsespoor@compuserve.com

Dr. Sami **Pihlström**
University of Helsinki
Department of Philosophy
P.O. Box 24, FIN-00014 University of Helsinki, Finland
sami.philstrom@helsinki.fi

Prof. Dr. Graham **Priest**
University of Melbourne
Department of Philosophy
Melbourne, Victoria 3010, Australia
g.priest@unimelb.edu.au

Prof. Dr. Edmund **Runggaldier**
Universität Innsbruck
Institut für Philosophie, Katholisch-Theologische Fakultät
Karl-Rahner-Platz 1, A-6020 Innsbruck
edmund.runggaldier@uibk.ac.at

Prof. Dr. Gerhard **Schurz**
Universität Erfurt
Lehrstuhl für Wissenschaftsphilosophie
Nordhäuser Str. 63, D-99089 Erfurt
gerhard.schurz@uni-erfurt.de

Prof. Dr. John R. **Searle**
University of California, Berkeley
Philosophy Department
314 Moses Hall, Berkeley, CA 94720-2390
searle@cogsci.berkeley.edu

Prof. Dr. Ulrich **Steinvorth**
Universität Hamburg
Philosophisches Seminar
Onckenstr. 26, D-22607 Hamburg
ulstein@philosophie.uni-hamburg.de

Prof. Dr. Frederik **Stjernfelt**
University of Copenhagen
Department of Comparative Literature
Njalsgade 80, DK-2300 S
stjern@coco.ihi.ku.dk

Prof. Dr. Avrum **Stroll**
Philosophy Department
University of California, San Diego
UCSD, 9500 Gilman Drive, La Jolla, California 92093
astroll@ucsd.edu

Prof. Dr. Barbara **Tversky**
Stanford University
Department of Psychology
Jordan Hall, Bldg. 420, Stanford, CA 94305-2130
bt@Psych.Stanford.EDU

Prof. Dr. Max **Urchs**
Universität Konstanz
Fachbereich Philosophie
PF 5560 D22, D-78434
Max.Urchs@uni-konstanz.de

Prof. Dr. Jan **Woleński**
Jagiellonian University, Krakow
Institute of Philosophy
Grdzka 52, 31-044 Krakow, Poland
wolenski@theta.uoks.uj.edu.pl

Schriftenreihe der Wittgenstein-Gesellschaft: Band / Volume 26

THE ROLE OF PRAGMATICS IN CONTEMPORARY PHILOSOPHY
Proceedings of the 20th International Wittgenstein-Symposium
DIE ROLLE DER PRAGMATIK IN DER GEGENWARTSPHILOSOPHIE
Akten des 20. Internationlen Wittgenstein-Symposiums

Kirchberg am Wechsel (Austria) 1997
Hrsg. / Eds. Paul Weingartner, Gerhard Schurz, Georg Dorn
Wien 1998, 416 Seiten, geb., ISBN 3-209-02585-1

How are scientific theories related to problems of practical application? Is objectivity in science something above and beyond practical interests, or is it determined by them? Questions of this kind are in the focus of pragmatic philosophy. This volume documents the most important pragmatic approaches in contemporary philosophy. By of its outstanding contributors and high quality of presentation this volume recommends itself as reader in contemporary pragmatic philosophy. An extensive introduction explains the basic concepts and summarizes the contents of the papers.

Wie hängen wissenschaftliche Theorien mit praktischen Anwendungsproblemen zusammen? Steht objektive Wissenschaft über allen praktischen Interessen, oder ist sie von solchen bestimmt? Derartige Fragen stehen im Zentrum pragmatischer Philosophie. Der vorliegende Sammelband dokumentiert die bedeutendsten pragmatischen Ansätze der Gegenwartsphilosophie. Hochrangige internationale Besetzung und didaktische Aufbereitung empfehlen diesen Band als Textbuch in pragmatischer Gegenwartsphilosophie. Eine ausführliche Einleitung erklärt die wesentlichen Grundbegriffe und faßt die Beiträge übersichtlich zusammen.

Table of Contents / Inhaltsverzeichnis

Preface / Vorwort
Gerhard Schurz: Introduction and Overview

I. Pragmatism and Pragmatics in Contemporary Philosophy
I. Pragmatismus und Pragmatik in der gegenwärtigen Philosophie
Nicholas Rescher: Pragmatism in Crisis
Gerhard Schurz: Kinds of Pragmatisms and Pragmatic Components of Knowledge
Peter H. Hare: Classical Pragmatism, Recent Naturalistic Theories of Representation, and Pragmatic Realism
Paul Gochet: Foundherentism and Pragmatism Revisited

II. The Role of Pragmatics for Language and Communication
II. Die Rolle der Pragmatik für Sprache und Kommunikation
Peter Gärdenfors: The Pragmatic Role of Modality in Natural Language
Hubert Haider: Form Follows Function Fails – as a Direct Explanation of Properties of Grammar
Georg Meggle: Regeltheoretische contra Intentionalistische Semantik?

III. Pragmatic Approaches to Meaning, Reference and Truth
III. Pragmatische Ansätze für Bedeutung, Referenz und Wahrheit
Franz von Kutschera: Pragmatische Sprachauffassung, Bedeutungen und semantische Antinomien
Paul Weingartner: Pragmatic Presuppositions of Tarski's Truth Condition
Alexander Hieke und Edgar Morscher: Pragmatische Widersprüche
Risto Hilpinen: Pragmatics and Pragmatism: On C.S. Peirce's Pragmatic Theory of Meaning
Avrum Stroll: Direct Reference and Fiction

IV. Pragmatical Topics in Logic, Reasoning and Decision
IV. Pragmatische Themen der Logik, des Schließens und der Entscheidung
Ernest W. Adams: The Utility of Truth and Probability
Brian Skyrms, Peter Woodruff and Gary D. Bell: Parametric Conditionals, Stalnaker Conditionals, Natural Families, and General Bayesian Conditionals: What is Needed for Decision Theory?
John L. Pollock: Degrees of Justification
Evgenij A. Sidorenko: Binary Relational Semantics of Entailment: Epistemic and Formal Aspects

V. Pragmatical Insights in Philosophy of Science and Epistemology
V. Pragmatische Einsichten in Wissenschafts- und Erkenntnistheorie
Patrick Suppes: Pragmatism in Physics
Peter Mittelstaedt: Pragmatische Aspekte einer universell gültigen Quantentheorie
William L. Harper: Measurement and Approximation: Newton's Inferences from Phenomena versus Glymour's Bootstrap Confirmation
Theo A.F. Kuipers: Pragmatic Aspects of Truth-Approximation
Gerhard Vollmer: Woran scheitern Theorien? Zum Gewicht von Erfolgsargumenten
Ken Gemes: Logical Content and Empirical Significance

VI. Ethics from a Pragmatical Point of View
VI. Ethik aus pragmatischer Perspektive
Dieter Birnbacher: Praktische Ethik als ethische Pragmatik
Wolfgang Lenzen: Der Wert des Lebens – Möglichkeit und Grenzen eines pragmatischen Ansatzes
Julian Nida-Rümelin: Subjective and Objective Reasons

VII. The Pragmatic Dimension in Wittgenstein
VII. Die pragmatische Dimension bei Wittgenstein
David Pears: Saying and Doing: The Pragmatic Aspect of Wittgenstein's Treatment of "I"
Luis Flores H.: Elements for a Pragmatics in Wittgenstein's Philosophische Untersuchungen
Pamela L. Dick: First-Person Truth

Verlag Hölder-Pichler-Tempsky
A-1096 Wien, Frankgasse 4, Postfach 127 Tel.(+43-1) 40136-0* / FAX (+43-1) 40136-185
E-Mail: office@oebvhpt.at, Internet: www.oebvhpt.at/

Schriftenreihe der Wittgenstein-Gesellschaft: Band / Volume 27

APPLIED ETHICS / ANGEWANDTE ETHIK
Proceedings of the 21st International Wittgenstein-Symposium
Akten des 21. Internationalen Wittgenstein-Symposiums

Kirchberg am Wechsel (Austria) 1998
Hrsg. / Eds. Peter Kampits, Anja Weiberg
Wien 1999, 398 Seiten, geb. ISBN 3-209-02829-X

Applied ethics is one of the most important topics of nowadays philosophy. The advancement of science and technology has spawned a plethora of new ethical questions in the fields of ecology, economics, politics as well as in technology and medicine. Furthermore feminism also brings a challenge for ethics as innovatory and new disciplines like genetics.

Angewandte Ethik zählt gegenwärtig zu den wichtigsten Themen der Philosophie. Die durch den wissenschaftlich-technischen Fortschritt entstandenen neuartigen Fragestellungen reichen von der Ökologie, Ökonomie und Politik bis zur Technik und Medizin. Darüber hinaus haben feministische Fragestellungen ebenso wie neue Forschungsgebiete wie etwa die Genetik viele ethische Implikationen.

Table of Contents / Inhaltsverzeichnis

Preface / Vorwort

I. Wittgenstein
Neno Bogdanov: Leo N. Tolstoi und Ludwig Wittgenstein oder das Ethische als Grundprinzip
Bektur Esenkulov: Ludwig Wittgenstein's Ideas in the Context of Modern Philosophical Trends Formation in Kyrgyzstan
Matthias Kross: „Wenn etwas gut ist, so ist es auch göttlich" Wittgensteins Satz über die Ethik
Melika Quelbani: Das Schicksal der Ethik in der Post-Tractatus-Philosophie
Goran Svob: Is Identity a Relation?
Jure Zovko: Platon und Wittgenstein – Ein Vergleich

II. Angewandte Ethik
II. Applied Ethics
Kurt Bayertz: Moral als Konstruktion Zur Selbstaufklärung der angewandten Ethik

III. Wirtschafts- und Berufsethik
III. Business Ethics
Franz Rupert Hrubi: Pyrrhussieg im Wettbewerb der Lebensformen?
Peter Koslowski: Shareholder Value als das Kontrollprinzip und nicht als der Zweck des Unternehmens
Angelika Krebs: Recht auf Arbeit oder Grundeinkommen?

IV. Politik und Ethik
IV. Politics and Ethics
Allen Buchanan: Humanitarian Intervention and the Natural Duty of Justice
Ridha Chennoufi: Ethik und Kolonialismus
Gerald Doppelt: Liberalism, Multiculturalism, and the Politics of Identity
Reinhard Kleinknecht: Gibt es objektive ethische Maßstäbe für politisches Handeln?
Julian Nida-Rümelin: The Idea of a Global Civil Society
Myroslav Popovych: The ethical dimension of a post-communist political life: Ukrainian experience
Hans-Martin Schönherr-Mann: Politische Philosophie und praktische Hermeneutik. Die Transformation des politischen Handelns in eine Gesprächskultur
Rudolf Steindl: Die Beeinflussung der Todesproblematik durch Veränderungen der politischen Strukturen in Ost-Mitteleuropa
Alan Thomas: Values, Secondary Qualities and the Challenge of "Non-Objectivism"

V. Technik und Ethik
V. Technics and Ethics
Christoph Hubig: Sachzwänge: Herausforderung oder Entlastung einer Technik- und Wirtschaftsethik?
Günther Ropohl: Welche Schwierigkeiten die Technik mit der Ethik hat

VI. Umweltethik
VI. Environmental Ethics
Kyohun Chin: Umweltethik aus der anthropologischen und ästhetischen Perspektive der koreanischen Tradition
Dietmar von der Pfordten: Welche Entitäten sind ethisch zu berücksichtigen?

VII. Bioethik und Medizinische Ethik
VII. Bioethics and Medical Ethics
James V. Bachman: Religious Voices in Secular Settings
Frances Kamm: Physician Assisted Suicide, Intended Death, and the Ground of Value
James Lindemann Nelson: Agency by Proxy
Hans-Martin Sass: Differentialethik. Über die notwendige Integration von Fakten und Normen in Medizin und Biowissenschaften
Harri Wettstein: Sterben in Würde

VIII. Genetik und Ethik
VIII. Genetics and Ethics
Dieter Birnbacher: Somatische Gentherapie – ethische Aspekte
Allen Buchanan: Ethical Issues in Genetics

IX. Feminismus und Ethik
IX. Feminism and Ethics
Hilde Lindemann Nelson: Resistance and Insubordination: A Feminist Response to Medical Hegemony
Martha Nussbaum: In Defense of Universal Values

Autorenverzeichnis
List of authors

öbv&hpt A-1090 Wien, Frankgasse 4, Tel.(+43-1) 401 36-0* / FAX (+43-1) 401 36-185
E-Mail: office@oebvhpt.at, Internet: www.oebvhpt.at

Schriftenreihe der Wittgenstein-Gesellschaft: Band / Volume 28

Metaphysics in the Post-Metaphysical Age
Metaphysik im postmetaphysischen Zeitalter
Proceedings of the 22nd International Wittgenstein Symposium
Akten des 22. Internationalen Wittgenstein-Symposiums

Kirchberg am Wechsel (Austria) 1999
Hrsg. / Ed. Uwe Meixner
Wien 2000, 368 Seiten, geb. ISBN 3-209-03194-0

Table of Contents / Inhaltsverzeichnis

1. Wittgenstein mit und ohne Metaphysik
Wittgenstein With and Without Metaphysics
Klaus Puhl, Geschichte und Ritual. Wittgenstein und Foucault über genetische Erklärungen kultureller Praktiken
Eike von Savigny, The Later Wittgenstein's Explanatory Metaphysics
Joachim Schulte, Wittgenstein's Quietism
Scott A. Shalkowski, Theory Neutrality and Logic

2. Logik, Sprache und Metaphysik /
Logic, Language, and Metaphysics
Peter van Inwagen, "It is Nonsensical to Speak of the Total Number of Objects"
Christian Kanzian, Das Klassische Kriterium für Änderungen
Tomis Kapitan, Indexical Metaphysics
Maria E. Reicher, Predicable and Non-Predicable Universals
Gonzalo Rodriguez-Pereyra, Truthmaking and the Slingshot
Mark Sainsbury, Knowing Meanings and Knowing Entities

3. Metaphysik der Modalität / Metaphysics of Modality
Graeme Forbes, Essentialism Reconsidered
Peter Forrest, Actuality, Consciousness and Freedom: A Metaphysics for a Post-Metaphysical Age
Wolfgang Malzkorn, On the Conditional Analysis of Dispositions
Ulrich Nortmann, Essentialistische Konditionale für Extensionalisten

4. Metaphysik und Wissenschaft /
Metaphysics and Science
William L. Craig, Naturalism and Cosmology
Phil Dowe, The Improbability of Time Travel
Jan Faye, Beyond Science
Renate Huber, Zur philosophischen Beurteilung physikalischer Theorien: Poincaré und Einstein
Ingvar Johansson, Presuppositions for Realist Interpretations of Vectors and Vector Addition
Bernulf Kanitscheider, Die Feinabstimmung des Universums – ein neues metaphysisches Rätsel?
Johanna Seibt, Processes in the Manifest and Scientific Image
Barry Smith / Achille C. Varzi, Environmental Metaphysics

5. Metaphysik, Epistemologie und Axiologie /
Metaphysics, Epistemology and Axiology
Winfried Löffler, Was ist eigentlich revisionäre Metaphysik?
Nicholas Rescher, Optimalism and Axiological Explanation in Metaphysics
Richard Schantz, Empirismus und Realismus

6. Metaphysik des Geistes / Metaphysics of Mind
Herbert Hochberg, A Simple Refutation of Mindless Materialism
Uwe Meixner, From Skepticism to Metaphysics, or: The Core of Dualism
Martine Nida-Rümelin, Personal Identity and Substance Dualism
Achim Stephan, Are We Still Trying to Square the Circle? – Almost Optimistic Remarks on the Philosophical Problem of Phenomenal Consciousness

7. Metaphysik und Philosophiegeschichte /
Metaphysics and History of Philosophy
Richard Gaskin, Die Einheit der Aussage
Michael-Thomas Liske, Einzelsubstanz versus Holismus
Edith Puster, Zur Frage der Auflösbarkeit des Cartesischen Zirkels
Hermann Weidemann, Freiheit als metaphysisches Problem in der Philosophie der Antike

8. Metaphysikkritik / Criticism of Metaphysics
Jaroslav Peregrin, Metaphysics as an Attempt to Have One's Cake and Eat It
Erwin Tegtmeier, Metaphysical Mistakes: Old and New

öbv&hpt A-1090 Wien, Frankgasse 4, Tel.(+43-1) 401 36-0* / FAX (+43-1) 401 36-185
E-Mail: office@oebvhpt.at, Internet: www.oebvhpt.at