Making Ranking Theory useful for Psychology of Reasoning

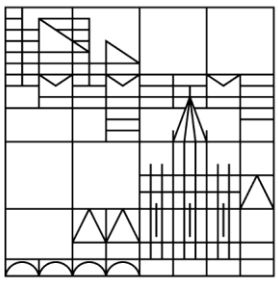
Niels Skovgaard Olsen

Making Ranking Theory useful for Psychology of Reasoning

Dissertation zur Erlangung des akademischen Grades
Doctor philosophiae (Dr.phil.)

an der

Universität Konstanz



Geisteswissenschaftliche Sektion

Fachbereich Philosophie

vorgelegt von Niels Skovgaard Olsen

Tag der mündlichen Prüfung: 17. November 2014

Referenten: Prof. Dr. Wolfgang Spohn
Prof. Dr. Sieghard Beller
Prof. Dr. Thomas Müller

Abstract

An organizing theme of the dissertation is the issue of how to make philosophical theories useful for scientific purposes. An argument for the contention is presented that it doesn't suffice merely to theoretically motivate one's theories, and make them compatible with existing data, but that philosophers having this aim should ideally contribute to identifying unique and hard to vary predictions of their theories.

This methodological recommendation is applied to the ranking-theoretic approach to conditionals, which emphasizes the epistemic relevance and the expression of reason relations as part of the semantics of the natural language conditional. As a first step, this approach is theoretically motivated in a comparative discussion of other alternatives in psychology of reasoning, like the suppositional theory of conditionals, and novel approaches to the problems of compositionality and accounting for the objective purport of indicative conditionals are presented.

In a second step, a formal model is formulated, which allows us to derive quantitative predictions from the ranking-theoretic approach, and it is investigated which novel avenues of empirical research that this model opens up for.

Finally, a treatment is given of the problem of logical omniscience as it concerns the issue of whether ranking theory (and other similar approaches) makes too idealized assumptions about rationality to allow for interesting applications in psychology of reasoning. Building on the work of Robert Brandom, a novel solution to this problem is presented, which both opens up for new perspectives in psychology of reasoning and appears to be capable of satisfying a range of constraints on bridge principles between logic and norms of reasoning, which would otherwise stand in a tension.

Zusammenfassung

Ein Leitmotiv dieser Dissertation ist die Fragestellung, wie man philosophische Theorien für empirische Wissenschaften nutzbar machen kann. Es wird ein Argument dafür aufgezeigt, dass es nicht genügt, seine Theorie theoretisch zu motivieren und sie mit bestehenden Befunden kompatibel zu machen, sondern dass man vielmehr dafür Sorge zu tragen hat, dass es möglich ist, Vorhersagen abzuleiten, die schwer zu variieren sind und von den bestehenden Theorien nicht geteilt werden.

Diese methodologische Empfehlung wird in Bezug auf den rang-theoretischen Ansatz zu Konditionalsätzen angewendet, welcher die epistemische Relevanz und das Ausdrücken von Gründen in der Semantik von Konditionalsätzen betont. In einem ersten Schritt wird dieser Ansatz gegenüber bestehenden Alternativen, wie etwa der suppositionellen Theorie von Konditionalsätzen, theoretisch motiviert. Dabei wird unter anderem ein neuer Lösungsansatz für das Problem der Kompositionalität und das Problem des objektiven Behauptungs-Charakters indikativer Konditionalsätze angeboten.

In einem zweiten Schritt wird ein mathematisches Modell formuliert, das uns erlaubt, quantitative Vorhersagen von dem rang-theoretischen Ansatz abzuleiten und es wird eingehend erörtert, welche neuen empirischen Fragestellungen damit verknüpft sind.

Abschließend wird das sogenannte Problem der logischen Omniszienz ausführlich behandelt, da es die Frage aufwirft, ob die Rangtheorie (und ähnliche Ansätze) auf zu idealisierten Rationalitätsannahmen beruht, um für die Psychologie des Denkens attraktiv zu sein. Aufbauend auf der Theorie Robert Brandoms wird dabei ein neuer Lösungsansatz angeboten. Dieser vermag es sowohl, neue Perspektiven für die Psychologie des Denkens zu eröffnen, als auch einer Menge scheinbar widerstreitender Forderungen, die an Brückenprinzipien zwischen Logik und Normen des Denkens gestellt werden, gerecht zu werden.

Table of Contents

Preface

The present dissertation is highly interdisciplinary in its nature. It was supervised by both a philosopher and a cognitive psychologist, and its list of references consists of about as much philosophy as psychology.

Doing interdisciplinary research is always challenging. Each discipline has its own traditions and internal standards. So there is a real danger of in attempting to meet the standards of two disciplines simultaneously, one succeeds in meeting neither. Moreover, when one's supervisors come from different disciplines, one is confronted with the related problem that the parts that are liked by the one are disliked by the other and *vice versa*.

In dealing with this difficulty, I have attempted to organize the dissertation in such a manner that there are chapters that are more densely philosophical (i.e. chapters I, II, and V) and others that are more of a psychological nature (i.e. chapters III and IV). However, as the topics are much intertwined, philosophical problems keep popping up in the psychological discussions and psychological issues keep emerging in the philosophical discussions. I have not tried to resist this tendency. On the contrary, I take it as a hallmark of the value of doing interdisciplinary research that the subject matter of either discipline can hardly be dealt with in isolation.

In a sense, the problem of how to do interdisciplinary philosophy is an organizing theme of this dissertation. These reflections arose out of a preoccupation with the question of how a researcher with a background in philosophy could make contributions to the interdisciplinary research project, *New Frameworks of Rationality*,[1] that empirical scientists would be capable of seeing the value of.

So to realize this project, chapter I is devoted to methodological reflections on how philosophers can make their theories useful for the empirical sciences (whenever this is indeed their goal, and they are theorizing about a subject matter that is indeed amenable to such use).

To illustrate the worth of these recommendations, chapters II-IV shift gears and consider a special case of a theory, which has been developed formally to an extent that it is ripe for application in experimental psychology; to wit, *the ranking-theoretic approach to conditionals* developed by my primary supervisor, Professor Dr. Wolfgang Spohn. In chapter II, this theory is motivated theoretically in a comparative discussion of the other prominent theories of conditionals currently finding application in psychology of reasoning. Here it is shown how we can theoretically motivate a relevance approach to conditionals in general, which ranking theory then provides a fruitful, formal explication of.

In chapter III, the first step of implementing the methodological recommendations from chapter I are taken, when a parallel between a statistical model called logistic regression and two-sided ranking functions is exploited to derive predictions from the ranking-theoretic approach to conditionals for a particular experimental paradigm in psychology of reasoning known as the conditional inference task. In chapter IV, a second step is taken, when it is considered to what extent the predictions derived in chapter III count as *unique* and *hard to vary*. Chapter IV moreover contributes to making the formal model of the conditional inference task from chapter III useful for experimental psychology by: (1) identifying its potential contributions to an existing theory of deductive and probabilistic reasoning called *the dual source approach* (which was developed by my secondary supervisor, Professor Dr. Sieghard Beller), (2) locating some of the predictions of the model within existing empirical findings, and (3) answering some psychologically motivated objections to the model.

In the final chapter, we shift gears for a second time, when the question is taken up of what to do about some of the idealizing assumptions of ranking theory in light of the recent rationality debates in psychology of reasoning. In particular, a strategy is proposed for how to make the assumptions of ranking theory more palatable giving rise to the problem of logical omniscience. The contribution of this chapter within the framework of the dissertation is then to increase the utility of making ranking theory accessible to experimental psychology by laying out a strategy for making its rationality assumptions less idealized. With it, this dissertation will not just have focused on the narrower topic of making the ranking-theoretic approach to conditionals accessible to experimental

psychology. But it will also have taken a first stab at engaging in the rationality debates currently taking place in psychology of reasoning. Philosophically this chapter is important, because an essential part of philosophy is an attempt of not getting lost in the details but maintaining a synoptic view of the larger picture. To this extent, it would have been unsatisfying for a dissertation with philosophical ambitions, if all it did was to make a part of ranking theory accessible to experimental psychology without reflecting on whether the assumptions of rationality embodied in it are too idealized to be applicable to real, psychological agents.

Moving to the formalities, an innovation in the present monograph is to use endnotes, which state the purpose of the note in headlines, so that the reader can make an assessment of which notes are worth the effort. All too often I find myself putting valuable points in footnotes only to discover in discussions that they haven't been read. To circumvent this problem, I will use endnotes (because they are more elegant) and mark their purpose (so that the reader is at liberty to be picky about what to read). I discourage skipping all the endnotes, because references, examples, and valuable information needed for the course of the argument will be contained in them (when I don't want to interrupt the flow of the argument). However, through the guidance of these headlines, it should be possible to skip a number of them.

Sections are referred to by following the convention that Roman numerals refer to chapters and Arabic numerals refer to sections of the chapter in question. Hence, section IV 2.2.2 refers to section 2.2.2 of chapter IV.

I am extremely grateful for all the support, discussions, and advice that I have received from my supervisors. Like all others who have contributed by their comments, dissent, or questions, I'll mark their influence at the beginning of the relevant chapters through endnotes. As will become apparent, I have also had the good fortune of having many colleagues, who were willing to comment on parts of the manuscript. I owe this circumstance in large part to the opportunities that *New Frameworks of Rationality* offered and to the stimulating philosophy community in Konstanz. In particular, I would like to thank Eric Raidl, Karl Christoph Klauer, Henrik Singmann, Michael De, Arno Goebel, Björn Meder, Igor Douven, Keith Stenning, Edouard Machery, Laura Martignon, Lars

Dänzer, and the members of Thomas Müller's colloquium for insightful and encouraging comments. Moreover, I should like to thank the members of reading groups on conditionals and ranking theory for good discussion.

A special thanks goes to David Over for encouraging me to work on establishing a connection between ranking theory and experimental psychology, which turned out to be a rewarding topic (even if it meant abandoning the direction, which the dissertation was originally taking). Furthermore, I thank friends and family, whose support has been invaluable. I dedicate this work to my parents Niels Viggo Skovgaard Olsen and Anne Grete Skovgaard Olsen and to my close friend Maria del Carmen Arana Flores.

Finally, I should add that as I deal with the preface paradox in chapter V, it is only appropriate, if I end my own preface by stating flat out that every claim made in the book appears to be true and justified to me. So for each particular claim in the book it holds that I am prepared to defend its correctness. Yet, I am also keenly aware that I do make mistakes some of the time. So if I were to assess my own work, I wouldn't want to presuppose that all the claims made in the book were correct. Indeed, if this book was written by somebody else, I wouldn't have made that assumption, so neither should you. (However, there would have been little point in writing it, if I didn't hope that you will find upon critical scrutiny that most of the assertions in the book are capable of functioning as a starting point for further inquiry.) To find out what to make of these paradoxical remarks, I am afraid that you will have to hold your breath until chapter V, dear reader.

[1] http://www.uni-giessen.de/cms/fbz/fb06/psychologie/abt/kognition/spp1516/

I

On how to make Philosophical Theories useful for Scientific Purposes[2]

Abstract: The purpose of this chapter is to come up with some concrete suggestions for how philosophers working in the broad area of cognitive science can help make their theories useful for scientific purposes. The main contention is that philosophers can help bridge this gap by themselves working out which predictions their theories are capable of generating. Two constraints that such predictions should be capable of satisfying are identified. The utility of these constraints is then illustrated using examples from psychology of concepts, and finally some general considerations are presented for how to generate predictions of philosophical theories in experimental psychology. In addition, an appendix has been added to vindicate the potentially controversial idea of expanding hypothesis spaces, which is relied on in the course of the argument.

1. Introduction

If philosophers want to make their theories useful for empirical disciplines, it is important to get clear about which criteria of adequacy their theories have to fulfill to serve scientific purposes. Using experimental psychology as our example, the goal of the present chapter is to throw some light on this issue.

I think that the short answer is that *unique* and *hard to vary predictions* should be identified and that the philosopher should contribute to providing *operationalizations* through the assignment of measurable quantities to the key theoretical notions (or constructs, as psychologists like to put it) whenever possible. The bulk of this chapter is spent on clarifying and justifying the first part of this claim by means of examples from psychology of concepts. It is of course clear that some division of labor has to be involved, since philosophers, without the professional training of experimental scientists, will probably neither be able to devise exact scientific models (e.g. computational and neurophysiological models) nor to make concrete plans for controlled experiments (and much less to actually carry out the experiments themselves). But they will still have made a contribution to the extent that they attempt to meet the criteria introduced above, or so I shall argue.

I say *attempt*, because providing good operationalizations of the kind of concepts that philosophers are interested in is in it itself something of a craft. If the operationalizations don't fully capture the phenomena under study in all their complexity, then this is one of the first things that critics will stumble upon,[3] and which operationalizations are made will determine the measurement scale that can be constructed, which in turn has implications for which statistical analyses can be applied to the data (Eid, Gollwitzer, & Schmitt, 2010).

Since many of the statistical techniques taught in standard psychology courses only apply to interval scales or higher (where different items are mapped into different numbers on a scale, which are ordered along a continuum with a fixed interval), it is probably preferable if the operationalizations enable the construction of such scales. If this assumption is violated (or the assumption is violated of the data being normally distributed or of the variance of the data points being homogenous), non-parametric statistics can be applied. However, non-parametric tests have a lower power (i.e. a lower chance of rejecting

the null hypothesis that there is no difference between the groups, when in fact there is a difference) *when the assumptions of parametric tests are fulfilled*.[4] And sometimes one may find oneself with an interesting research question for which no suitable non-parametric test has been developed (Howell, 1997: ch. 18, A. Field, 2009: ch. 5, 15). For these reasons, I assume that there is a preference for operationalizations on interval scales, ratio scales, or absolute scales.

Before we move on to the actual substance of this chapter, two important qualifications should be added. The first is that what I present here is put forward as a *hypothetical imperative* (i.e. "given that a philosopher has the goal of making his or her work useful for cognitive science, then this is what he needs to do") and not as a *categorical imperative* (i.e. stating what the philosopher must do under all circumstances). If the philosopher prefers to withdraw to his or her pure investigations, then this chapter will not try to make a case that it is best to do otherwise.[5] However, it is at the same time clear that such an attitude cannot be adopted in an interdisciplinary work such as the present. Hence, for our purposes there is no way around engaging with the methodological issue raised by the title of this chapter.

A second qualification is that by articulating the recommendation that philosophers should contribute to identifying unique and hard to vary predictions of their theories, the claim is not advanced that this is the *only* contribution that philosophers are capable of making to the empirical sciences. There are, of course, many other valuable contributions like the use of thought experiments, increasing conceptual clarity, and contributing to the interpretation of the results.[6] Rather the point is to give recommendations for the specific case of whenever philosophers make theories about some phenomenon that is within the reach of empirical investigation (which is something that happens more often than might be expected).

2. The Uniqueness Constraint

When philosophers attempt to make their theories useful for scientific purposes often most of the effort is invested in theoretically motivating them through a comparative discussion of the theories currently under consideration. Such efforts, if successful, establish that the theories in question have a high prior probability and that they are worthy of further investigation. Moreover, sometimes one will also see extensive attempts of showing that the theories are compatible with existing findings. Some recent examples include Varela, Thompson, & Rosch (1991), Gallagher & Zahavi (2008), and Andrews (2012).

However, such efforts do not yet suffice to show that the theories in question are also capable of laying the basis for a fruitful, empirical research program that is in a position to make novel discoveries of its own (as opposed to merely redescribing existing findings in a novel vocabulary). Accordingly, one rarely finds a focus on the fact that expanding the hypothesis space by further candidates can itself have detrimental effects as highlighted by the following argument:

(P_1) The goal of the empirical sciences is to reach an empirically grounded decision among the candidates under consideration.[7]

(P_2) Introducing more theoretical possibilities into the empirical discourse that are merely compatible with the existing data increases the uncertainty by leaving us with more possibilities that we don't know how to exclude again.

(C) Hence, philosophers do not contribute to the goal of the empirical sciences, if their input to the empirical discourse only consists in introducing more theoretical possibilities that are merely compatible with the existing data.

The general point of the argument is that if philosophers want to make their theories useful for scientific purposes (i.e. if they want scientists to view their theories as relevant for their activities), they must show how their theories actually help the scientists achieving something that they were already trying to achieve.

In this context, a problem arises for theories that are merely made compatible with existing findings. For not only has one failed to bring us closer towards the goal of finding out what the answer is by coming up with a theory that lacks unique predictions of its own. But one has actually also brought us further away from determining what the right answer might be to the extent that our now expanded hypothesis space contains more possibilities that we cannot experimentally discriminate between. So in a sense we are farther away from reaching the goal of our inquiry, if this is all the new hypothesis contributes with.

In philosophy, it is common to make contributions that merely consist in pointing out that there are theoretical possibilities that have previously been overlooked. Accordingly, the discussions can take many epicycles to the frustration of outside observers, who are under the impression that the discussions don't seem to be getting anywhere.[8] And on the top of that the constant challenging of basic assumptions of even successful theories, and the apparent overproduction of speculative alternatives that are kept alive for longer than in most other disciplines, may appear to be quite counterproductive.

Of course, this impression is not entirely adequate. For as I have argued in Olsen (2014) there is much progress in terms of gaining a better overview over the landscape of theoretical possibilities and discovering which explanatory challenges they are either capable or incapable of meeting. Moreover, although it may not be conducive to the generation of universal research programs, this way of organizing a discourse does have the advantage of enabling discoveries from unexpected avenues and subjecting the theories currently in vogue to a ruthless competition, whereby they stand under a constant pressure to be rethought and refined.

Nevertheless, the argument above is meant to highlight a danger that arises once participants of a discourse with roughly this shape start making contributions to a discourse organized around the different goal of reaching an empirically grounded decision among the candidates under consideration. Then they might find themselves inclined to doubt that the evidence speaks in favor of the current theories and populate the hypothesis space with further alternatives, which in turn increases the uncertainty. If we are to prevent this temporary drawback from becoming permanent, it is the burden of the philosopher to

show that the additional uncertainty introduced in the short run is outweighed by long-term benefits.

One way of doing so is by showing that the theory introduced is capable of generating *unique* predictions (i.e. predictions not shared by any of the other candidates in the restricted hypothesis space under consideration). The reason is that such predictions have the attractive features of either enabling us to exclude the possibility that the philosopher introduces (if the predictions are unsuccessful) or of strongly confirming it (if they are successful),[9] which in turn would contribute to excluding the other possibilities. Moreover, since these predictions are not shared by the other candidates, their formulation advances empirical research by leading to the formulation of new research questions. Here the temporarily increased uncertainty can in other words be tolerated, because we know in principle how to experimentally distinguish the new theory from the alternatives. If we are lucky, then these new predictions can lead to new discoveries. If not, it merely contributes to setting our temporarily increased uncertainty back to zero again.

Yet, a theory cannot just consist of unique predictions, if this comes at the cost of not being able to explain existing data. So the proper way of formulating the present constraint is that it is an attractive feature of a theory that it is able to generate unique predictions even if such predictions cannot stand alone.

Finally, to encounter the potential worry that a principled underdetermination by the empirical evidence would take the steam out of this proposal, I would like to end this section by making the following suggestions. In principle, Duhem and Quine may be right that hypotheses are not tested in isolation and that it may be possible to identify various sources of error of a failed experiment ranging from mistaken background assumptions, assumptions about how the measurement works, auxiliary hypotheses, the actual hypothesis itself etc. (cf. Stanford, 2013).

However, such worries need to be fought out in relation to specific experiments and for each individual experiment it needs to be made plausible that any of these alternative sources of error was actually a likely source of error in the given case. Furthermore, it seems that there is a tendency in the empirical discourse not to accept an alternative interpretation of the data until the point, where its own empirical fruitfulness can in turn be

established through the support of further experiments that introduce the additional control conditions and experimental manipulations that this interpretation would require. So each possible source of error is not on an equal standing in any given case. Some of them will remain speculations whose credentials cannot be established empirically. That is, an explanation of a pattern in the data is first truly convincing, when it can be shown that it not only accounts for what has been found, but that it identifies factors that can be varied to alter the pattern in predictable ways or even reverse it. And for some of the alternative interpretations one can come up with, this will simply not be possible.

Moreover, it should be noticed that the claims advanced above do not commit us to the view that the heuristic value of predictions that are not shared by any of the other candidates under considerations can be established by any single, decisive experiment. Rather it suffices that it be established in the long run. So if the expectations are disappointed about the outcome of a particular experiment, the results are replicated, various suggestions for what might have gone wrong are controlled for, and the predictions of the theory still fail to be satisfied, pressure is gradually built up towards the conclusion that the fault may very well lie in the theory itself. As a result, the attempts of its proponents to argue otherwise will eventually become more and more untenable. So even if the exclusion of possibilities cannot be established by a single, decisive experiment, the identification of unique predictions will still serve a heuristic function by pinpointing the junctures, where the battles need to be fought out.

In section 4 we will encounter examples from psychology of concepts illustrating the usefulness of this criterion.

3. Hard to Vary Predictions

The request for *hard to vary predictions* is due to the physicist David Deutsch (2011: ch. 1), who argues that a good explanation is one that is hard to vary while still accounting for what it purports to account for. As this indicates, Deutsch talks about hard to vary *explanations*, and not directly about hard to vary *predictions*, but the connection between the two will soon become clear.

Deutsch argues that what the explanation asserts about reality should be tightly constrained by the nature of explanandum. That is, the details of the explanatory mechanism that the explanation invokes should be such that they play a functional role in producing the explanandum. The idea is that when we ask why the explanation makes a particular assertion about reality as opposed to another, the answer should be framed in terms of what is required to account for what the explanation purports to explain, so that we are not at liberty to change parts of the explanation to dodge unpleasant challenges. Whenever this is the case, the theory cannot easily be adjusted to produce other predictions and often it will end up having a range of implications for other phenomena for which the theory was not originally designed.

In Deutsch's example, the axis-tilt theory of seasonal change[10] not only explains why seasons change, but it also explains why they are out of phase in the two hemispheres, why tropical regions do not have them, and why the summer sun shines at midnight in polar regions. Moreover, the theory implies that there must be seasonal variation on planets in other solar systems that equally have a tilted axis relative to their orbital plane. In the absence of prior knowledge about such phenomena, most of them are quite surprising. However, the originator of this theory would be hard pressed to make changes to the theory that would allow him to get rid of these implications if he wanted, because they all follow from the explanatory mechanism the theory invokes given some basic physical and geometrical facts. As this example makes clear, the fact that the details of the theory's explanatory mechanism are so tightly constrained ensures that it will hold for some of its predictions that they cannot easily be varied without the theory ceasing to account for what it purports to account for. It is this feature of the kind of theories in question (i.e. their ability to produce such hard to vary predictions), which makes us capable of learning something from testing them.

Deutsch illustrates this latter point by contrasting the axis-tilt theory of seasonal change with pre-scientific myths (e.g. that the periodic sadness of the goddess of earth and agriculture explains the arrival of winter) and arguing that the problem with the latter is not so much that they don't produce *any* testable predictions at all, as it is commonly thought (after all, the arrival of winter should be uniform, if it is merely regulated by when the

sadness of this goddess sets in). But the problem is rather that the explanatory mechanism invoked could easily have been varied to produce other predictions, insofar as there is no particular functional reason for why exactly it is the periodic sadness of this specific goddess—as opposed to some other antics by the ancient gods—that explains the arrival of winter. So even when we do discover the falsity of particular predictions that such theories make, we have not progressed towards identifying possible errors that may be lurking in our theories, since these theories could just be adjusted in an *ad hoc* manner to produce whatever other prediction that fits for the occasion.

Rather some of the predictions of a theory should be based on its core claims and count as being beyond repair by auxiliary hypotheses. The point is that only by testing predictions, which cannot easily be varied without the theory ceasing to account for what it purports to account for, will we be in a position to reduce our uncertainty about which theory to accept. Since this holds regardless of the uniqueness of our predictions, both of our constraints have to be imposed together in the sense that successful predictions that have either one of these features are to be weighed more than those that lack them. Their justification thus ultimately rests on their ability to help us trim down the number of serious possibilities in our hypothesis space.

At this point further possible constraints (like parsimoniousness in the number of free parameters that have to be estimated on the basis of the data) could be considered, and one would thereby find oneself entangled in difficult decisions about how to weigh these various considerations. However, our concern here is not so much with an exhaustive discussion but rather with the identification of some useful criteria, which can easily be applied. So to encounter the potential worry that it may not be possible to reach Deutsch's stern requirement in many other disciplines than perhaps in physics, I will now illustrate its utility by drawing on examples from psychology of concepts.

4. Examples from Psychology of Concepts

For the purposes of this discussion, I will focus on the classical view of concepts, the prototype theory, and the exemplar view (thus temporarily bracketing other psychological

theories, such as the knowledge approach, for simplicity). In doing so, I will take the presentations of the state of the art provided by Murphy (2004) and Machery (2009) for granted and merely show how the criteria from sections 2-3 can be applied to some of their arguments in order to illustrate their utility.

But before we begin, it is useful to keep in mind that the notion of concepts that plays a role in psychology can be characterized as consisting in information that is stored in long-term memory, which is used by default in the cognitive processes underlying higher cognitive competences such as categorization, deduction, induction, analogy-making, linguistic understanding, and planning (Machery, 2009: ch. 1). As a result, psychological theories of concepts are required to characterize the kind of information that is stored in concepts, how it is acquired, and the kind of mechanisms that utilize it.

A good example of how philosophical theories can find application in psychology is *the classical view of concepts*,[11] which holds that the nature of concepts consists in definitions specifying necessary and sufficient conditions for category membership. As part of adopting this philosophical theory for psychological purposes, it was added that the mechanism underlying categorization consists in comparing candidate entities with the properties outlined in the definition, which in turn had to be mentally represented somehow, in order to see whether they fulfilled the given properties.

Simple as it may be, this theory is already ripe with implications: (1) membership requires that the object satisfies all the conditions outlined in the definition, (2) every object is either a member of the category or not a member; there can be no intermediates, (3) for all the members of the category it holds that they are equally good members (and *vice versa* for all the non-members), and (4) the fact that categories can be hierarchically ordered can be explained by the transitivity of category membership that follows from the possibility of including the definition of one category (e.g. mammal) in the definition of another (e.g. whale).[12]

So if the psychological theory based on the classical view is supposed to have any purchase on specifying our understanding of concepts, it is to be expected that subjects are sensitive to these properties. The pressing question for us to consider is then whether the predictions that can be based on this are unique and hard to vary. Compared to the two

other alternatives that we shall look at, these predictions are certainly unique (in fact, part of what made these other theories so popular was precisely that they differed from the classical view on this score, because they were thereby better able to account for the empirical data).[13]

But are they also hard to vary? The mere fact that there exist at least two other versions of the classical view, which were designed with the explicit goal of not being committed to these particular predictions, might encourage us to think otherwise. But I will nevertheless argue that such a conclusion would be premature.

First, we need, however, to take a look at some of the empirical findings that caused a problem for the classical view. To be sure, the classical view is also beset by theoretical difficulties, which are well-known in philosophy, such as: (i) problems arising from Sorites Paradox and the fuzziness of most of our ordinary concepts (van Deemter, 2012), and (ii) the problem of coming up with good definitions for even the simplest of them, as emphasized by the late Wittgenstein. Among the empirical findings, other results have been reported as well, but here we just focus on what is known as *the typicality effect*, which historically played a large role in overturning the classical view:

> *The Typicality Effect*: members of categories that are rated to be typical (e.g. the fruit orange) are categorized more quickly and accurately than members of categories that are rated as atypical (e.g. the fruit olive). How typical a member of a category is rated to be moreover influences many other cognitive processes involving conceptual content. Examples include: 1) subjects rarely change their minds about category judgments involving typical members (in contrast to those involving atypical ones), 2) typical members are more often produced, when subjects are asked to name instances of a category, 3) encountering typical members of a category facilitates concept learning, 4) typicality effects in non-monotonic reasoning,[14] and 5) typicality effects in category-based inductions.[15] (Murphy 2004: ch. 2, Machery, 2009: ch. 6)

The problem for the classical view is that it lacks a way of distinguishing between typical and atypical members of categories, because if an entity meets all the conditions outlined in the definition it should be just as good a member as any other. So due to the

way category membership is specified by this theory, there are good functional reasons for why we shouldn't expect to find the typicality effect at all. Yet, as Murphy (2004: 22) points out: "Typicality differences are probably the strongest and most reliable effects in the categorization literature".

In contrast, the typicality effect is a hard to vary prediction of *the prototype theory*, which historically led to its discovery. The reason is that according to the prototype theory, the mechanism underlying categorization consists in comparing candidate entities for their similarity with a weighted summary representation containing statistical knowledge about the distribution of properties in a given category.[16] If they reach a certain threshold of similarity, they are judged to be members of the category. As a result, the typicality effect in categorization is to be expected, because this threshold is more easily reached in the case of typical members (given that these are exactly the subset of the category's extension that have many of the highly weighted properties, which implies that less features have to be considered before the threshold is reached).

To take another example, if concept learning consists in forming a weighted summary representation of the entire category, it should be facilitated by encountering typical members to the extent that these are exactly the members of the category that have many of the properties shared by a large portion of the other members. Of course, the prototype theory is capable of generating other predictions as well. But these few examples already illustrate the point we need to make: it would have posed a serious challenge for the prototype theory, if the typicality effect hadn't been found, because it is hard to see how a theory could possibly invoke cognitive processes of the kind described, without positing an advantage in information processing for the typical members of a category. It is for this reason that we presented the typicality effect as a hard to vary prediction of the prototype theory above.

The revised version of the classical view tries to accommodate the findings mentioned by positing that concepts consist of two components: (i) an identification procedure that is used for a first, quick categorization, which essentially consists of the kind of conceptual content emphasized by the prototype theory (i.e. characteristic features like the fur of dogs, which are not themselves definitional, but which are nevertheless useful for identifying

members of the category), and (ii) a core that consists of the definition of the concept, which is used for more careful categorization tasks. This revised version in other words involves a compromise; on the one hand, it builds into the classical view a component that doesn't generate any unique predictions of its own to account for the apparently incompatible data, and on the other, it retains definitions as the core of concepts, which gives it the potential for producing unique predictions.

Unfortunately for this theory, it turns out that the latter component is not really needed to account for any of the data. So we are left in the paradoxical situation that what was supposed to be *the core* of concepts is apparently unneeded to account for most of the roles that concepts play in our mental lives. As Murphy (2004: 28) says:

> almost every conceptual task has shown that there are unclear examples and variation in typicality of category members. Because the concept core does not allow such variation, all these tasks must be explained primarily by reference to the identification procedure and characteristic features.

Accordingly, the theory is now being rejected by the majority of psychologists working in the field (ibid.).

But if there exists a revised version of the classical view that is compatible with the typicality effect, why did we then say above that the predictions that go against it were hard to vary predictions of the original version? The reason is that although the original version and the revised version agree on there being a definitional core of our concepts, the two versions are actually quite different as psychological theories. Not only do they posit different conceptual contents to be stored in long-term memory, but they also invoke different mechanisms to explain categorization. However, if predictions are only hard to vary relative to a particular explanatory mechanism, and the two theories differ in this respect, it is hardly surprising that these two theories don't share the same hard to vary predictions. So for our purposes, the original version of the classical view and the revised version should really be considered as distinct psychological theories, even if they do share a certain family resemblance.

A second attempt to revise the classical theory is to be found in Rey (1983), which is named *the external classical view*. Inspired by the work on proper names and natural kind terms by Putnam and Kripke, Rey insists on a stark opposition between epistemological issues, having to do with our procedures for identifying the referents of our concepts, and metaphysical issues, dealing with their identity conditions. Whereas psychology of concepts has plenty to say about the former, Rey's point is that its discoveries are silent on the latter issue. The idea is then that the classical view might be defended from the criticism based on apparent incompatible data, if definitions are taken as specifying the identity conditions of concepts, regardless of whether the latter are ever known to the participants, because what counts is only what would ultimately—under some idealized conditions—be used to decide whether a candidate meets the conditions. In an attempt to give this theory a psychological dimension, Rey articulates the idea of our concepts possessing empty slots for definitions, which can later be filled out by the relevant sciences, just as the atomic number of gold may now be taken to provide an adequate definition of this concept.

However, aside for whatever illumination this suggestion brings about our common sense metaphysics, the otherwise imaginative psychologists Smith, Medin, and Rips (1984), must admit that they are hard pressed to see how this idea of slots for definitions in our concepts, which may never be filled, can generate any useful predictions for psychology of concepts. Furthermore, it is also hard to see from an evolutionary point of view why we should suppose that the brain would waste storage capacities for such empty slots that may in principle never influence the actual computational processes. Yet, in all fairness, it should be mentioned that Rey has tried to amend the situation in subsequent writings by attempting to spell out some possible predictions[17] and that the developmental psychologist Carey (2009: ch. 13) is more optimistic about this type of approach. But what interests us here is not so much Rey's fully developed position and whether there are in the end ways of saving it from criticism. So what we are going to do now is intentionally to create a straw man, who only wrote the original paper, and use this straw man to illustrate one of our basic methodological lessons.

The point is that whereas the revised classical view still tried to make a contribution to psychology of concepts, the strategy of our straw man is to find a loophole that allows

him to maintain definitions as the core of concepts by insisting that they are part of a metaphysical theory that holds irrespectively of whether such definitions will ever be known to the subjects possessing the concepts in question. As such, our straw man is left with a philosophical theory about the nature of concepts, which apparently lacks any substantial predictions for psychology of concepts—let alone any unique and hard to vary ones. So according to the argument in section 2, our straw man is guilty of not only not making much of a contribution to this field (regardless of whatever merits it may have as a philosophical theory), but also of introducing more uncertainty, without any immediate prospect of reaching an empirically grounded decision among the now increased number of theoretical possibilities.

Murphy (2004: 39-40) moreover argues that:

> In fact, much of the support such writers [philosophers and psychologists attempting to resurrect the classical view] give *for* the classical view is simply criticism of the evidence *against* it. For example, it has been argued that typicality is not necessarily inconsistent with a classical concept for various reasons, or that our inability to think of definitions is not really a problem. However, even if these arguments were true (and I don't think they are), this is a far cry from actually explaining the evidence. A theory should not be held just because the criticism of it can be argued against—the theory must itself provide a compelling account of the data. People who held the classical view in 1972 certainly did not predict the results of Rips et al. (1973), Rosch (1975), Rosch and Mervis (1975), Hampton (1979, 1988b, or 1995), or any of the other experiments that helped overturn it. It was only after such data were well established that classical theorists proposed reasons for why these results might not pose problems for an updated classical view.

So Murphy's point is that neither of the revised versions that we have looked at can really be considered serious contenders in psychology of concepts at this stage, because the classical view had no stake in making some of the most important discoveries in the field and apparently cannot come up with any explanation of its own of the pertinent results—even if it can be modified to become compatible with them after the fact, when all the important discoveries have already been made.

If it wasn't clear already, this strongly suggests that if philosophers want to make their theories useful to empirical disciplines such as psychology, they also have to contribute to making some of the important discoveries. It is mainly for this reason that the present chapter puts such a large focus on the need for philosophers occupied with cognitive science to start working out possible predictions of their theories.

Yet, we did identify unique and hard to vary predictions of the classical view in the form in which it was originally introduced as a psychological theory. So according to our own criteria, this version must be considered an excellent contribution to experimental psychology. By providing unique and hard to vary predictions, it was able to contribute to progress in the field by giving experimental scientists a clear hypothesis that they could experimentally contrast with other alternatives. Of course, the theory is now more or less universally rejected by the scientists in question. But this shouldn't blind us to the fact that learning about its problems was an important discovery that moved psychology of concepts forward. Measured in terms of progress, a demonstrably false theory that is recognized as such conveys a lot of information, especially if it has as much initial plausibility as this particular one did.

But notice that a whole-sale rejection of the classical view might also be mistaken, as the psychologist Smith (1989: 60) points out:

> This [that there are almost no concepts apart from bachelor for which definitions are available] is simply not the case. Classical concepts abound in formal systems that many people have knowledge about: consider *square, circle,* etc. from the system of geometry; *odd number, even number,* etc. from various systems of mathematics; *robber, felon,* etc. from the legal system; *uncle, nephew,* etc. from the kinship system; *island, volcano,* etc. from the geological system; and so on. In addition, many social concepts may have a classical structure. Consider concepts about national origin, such as *German* and *Italian,* where the core seems to come down to a few defining properties (e.g. either born in Germany or adopted citizenship in Germany). A similar story may hold for concepts of race (e.g. *Black, White),* gender *(male, female),* and profession (e.g. *lawyer, baker).* These social concepts are among the most widely used in categorizing other people. So there are plenty of cases of classical concepts, certainly enough to take seriously the idea that they constitute an important type of concept.

The point is that the classical view seems especially suited for concepts in stipulated formal systems.[18] This is perhaps not surprising considering, on the one hand, that the classical view goes hand in hand with classical logic both in virtue of their shared commitment to the law of excluded middle and of the contribution that necessary and sufficient conditions of concept-application make to a truth-functional decomposition of sentences (cf. Murphy 2004: ch. 2). And on the other, that if this theory of concepts has been the prevailing one for most of our recent, intellectual history, it would quite naturally serve as a guideline when new concepts were explicitly introduced—regardless of the consequent lack of commonality with all other concepts.

But be that as it may, we still have to consider the final contender of our short exposition. It was said that the typicality effect is a hard to vary prediction of the prototype view, but the prediction is actually not unique to it (even if we disregard the revised classical view, which in effect just introduces a component that copies the explanation that other theories give of this phenomenon). The reason is that there is another theory called *the exemplar view*, which also has the typicality effect as a natural consequence. According to this theory, the information associated with a concept in long-term memory, which is activated in conceptual tasks, consists of a set of memory traces of the individual members of the category. Consequently, the mechanism underlying categorization is taken to consist in a comparison based on similarity of the candidate entity with the whole set of remembered members, rather than with a unified, summary representation that contains statistical information about the distribution of properties within the category.

Moreover, concept learning is now taken to consist in forming memory traces of exemplars encoded as members of the category, instead of involving the formation of an abstract, summary representation of the entire category. Crucially, the exemplar view also predicts the typicality effect due to the fact that the typical members are similar to many exemplars of the category, which in turn implies that it should be easier to find evidence of membership when these entities are compared to memory traces of the exemplars stored in long-term memory compared to when atypical members are.

Interestingly, Machery (2009: 172) points out in this context that it was the received view throughout the 1980s and 1990s that the typicality effect couldn't be used to make a

decision between the two theories (in spite of the fact that further developments, which Machery also discusses, have changed the situation in this respect). This example thus once more highlights the importance of identifying unique predictions of one's theory in order to avoid the danger of a stalemate, where the pertinent experiments lose their ability to act as a tribunal over competing theories. Hence, the possibility of such a stalemate suggests that it is not always sufficient for the participants of a scientific discourse merely to do everything they can to attack the arguments of their adversaries, and to defend their own theories against criticism, as long as no unique predictions are identified. With this observation ends the illustration of the utility of the criteria from section 2 and 3 for experimental psychology.

5. Generating Predictions

To the extent that the preceding argument has been successful, it should at least have been made plausible that if philosophers want to make contributions to disciplines such as experimental psychology, they need to put some effort into identifying unique and hard to vary predictions of their theories. What I want to do now is to briefly raise some issues about how this can be done in relation to experimental psychology.

In philosophy it is common to remain satisfied with having analyzed some cognitive competence in terms of the activity of mental faculties (such as a faculty of reason or of judgment) without worrying about the underlying cognitive processes,[19] in spite of the fact that one is thereby depriving oneself from an important source of predictions in psychology, as illustrated by the discussion of psychology of concepts in section 4.[20]

In this context, Mareschal (2010) has usefully suggested that the kind of questions that one should provide answers to in order to identify the predictions of a theory in psychology are questions like: (i) how do things break down/fail, (ii) how can behavior be modified, and (iii) how can failures be rectified. And, of course, all of these are questions that are best answered in the light of concrete hypotheses about the underlying cognitive processes.

However, philosophers face a severe challenge when attempting to contribute to the latter enterprise due to the fact that they thereby have to restrain themselves from relying on popular philosophical methodologies involving the use of introspection and self-reports assisted by commonsense. The reason is that there is little reason to believe that anything other than the outputs of these cognitive processes are accessible to consciousness, as it is often pointed out in the psychological literature.[21] In other words, the philosopher, who wants to specify concrete predictions of his or her theory in this domain, runs into the problem of what source for hypotheses about underlying cognitive processes he or she is to rely on, now that the use of the abovementioned methodologies has been banned.

It seems that there is no way around engaging with the relevant empirical literature, and that a further remedy suggests itself. Throughout history aspects of the human mind have been compared to everything from a steam engine, to a telephone, a computer, and more recently to the apps on a smart phone. The proposed remedy is that one should accept such analogies as a source of well-understood models through their comparisons of unknown cognitive processes with the workings of devices that we know how to manipulate and construct, instead of just dismissing them by pointing out the obvious disanalogies, as it is done in quotes like the following:

> The brain is no more a computer than it is a central telephone exchange (the previously favoured analogy), and the mind is no more a computer programmer than it is a telephonist. (Bennett & Hacker, 2003: 65)

> The suggestion that we should think of ourselves as computer programs is not coherent. Human beings are animals of a certain kind. They weigh so-and-so many kilograms, are of such-and-such a height, are either male or female; they are born, grow, fall in love, get married and have children, and so forth – none of which can intelligibly be said of computer programs. (ibid: 432).

The reason why we are in need of such simplified models already emerges out of Deutsch's discussion: one of the things that makes the axis-tilt theory a better explanation of seasonal change than the pre-scientific myths is that we already know from examples on

a small scale how a surface is heated less when it is tilted away from a heat-radiating body than when tilted towards it, whereas we have no model for understanding the alleged relationship, whereby the periodic sadness of a goddess is able to set off the emergence of winter.

But a general analogy is not yet a concrete, testable model as Gigerenzer (1988) points out. His rich discussion moreover illustrates: (1) how dealing with interpretational problems in analogies exploited for the purpose of formulating new theories can show that they are compatible with several competing models, (2) that they may come with blind spots and illusions that have to be corrected, and (3) how logical problems in the analogies can carry over as structural problems of the resulting theories. To this extent, analytical work on the underlying analogies of the kind that Hacker & Bennett (2003) try to deliver is to be welcomed, *if it is done in the service of coming up with better models of the underlying cognitive processes*. Yet, this constructive side of the use of analogies as a platform for formulating new scientific models, which Gigerenzer (1988) so insightfully discusses, tends to go missing in quotes like the ones cited above.

In addition to the examples that Murray & Gigerenzer (1987) and Gigerenzer (1988) discuss, examples of such a constructive use include: (i) how Johnson-Laird (2008) based on a general analogy between computers and cognition uses computer programs that he writes as part of the activity of formulating concrete hypotheses about how *the mental models* of premises are constructed, which his theory posits to account for deductive reasoning, (ii) how Gallistel & King (2010) using the same analogy introduce constraints on effective computation known from computer science—like the use of a digital code, a ban on look-up tables, the need for the representation of variables, and the requirement of a symbolic read/write memory—to the cognitive processes in the brain, and (iii) Konrad Lorenz and Sigmund Freud's use of an analogy between aggression and hydraulic processes according to which a release of energy that instinctively builds up is the cause of aggressive behavior. What remained a metaphor in Freud's work has since been formulated as a scientific model capable of generating predictions known as the frustration-aggression theory. Of course, the analogy between aggression and hydraulic processes has subsequently fallen in popularity due to the fact that the assumption of aggressive energy is unsupported by

physiological evidence, and the frustration-aggression theory is now also rejected on independent empirical grounds.[22] But as we have already seen, even demonstrably false theories that are recognized as such can contribute to progress given that they have some initial plausibility. So if analogies with the newest pieces of technology can lead to the formulation of models that generate unique and hard to vary predictions, I don't see why philosophers should be against them. And if they are, the result is just to deprive themselves of an important source of predictions, which the methodological challenge mentioned above suggests that they can't really afford, whenever the goal is to make a contribution to empirical disciplines such as cognitive science.

Appendix 1:
Sampling Spaces and Prior Probability Distributions

In section 2, we encountered the idea of *adding* hypotheses to the hypothesis space. As this idea stands in contrast to the assumptions involved in applying Bayes' theorem and Shannon's theory of information, the purpose of this appendix is to challenge these assumptions.

Briefly stated,[23] Shannon's theory of information holds that communication of information requires that both the receiver and the transmitter have a representation of the set of all possible messages as well as a probability distribution over the messages within that set. The (average) information communicated by a given signal is then quantified in terms of how much the receiver can reduce its uncertainty in regard to the message (on average) when comparing before and after the signal was received relative to the receiver's probability distribution over the set of possible messages. The amount of expected uncertainty about the possible messages is called *Entropy* and it is to be calculated by the following formula:

$$H(X) = \sum_{i=1}^{n} P(x_i) \log_2 \frac{1}{P(x_i)} \quad (1)$$

Where $P(x_i)$ is the probability of the i^{th} message, or event, and the amount of information conveyed by it is given by:

$$\log_2 \frac{1}{P(x_i)} \quad (2)$$

Bayes' theorem (and other comparable conditionalization principles) can then be thought of as a rule specifying how the receiver should update his probability distribution over the set of possible messages as a function of the evidence gathered:

$$P(H|E) = \frac{P(E|H) \cdot P(H)}{P(E)} \tag{3}$$

The theorem states that the posterior probability of a hypothesis is to be determined by multiplying the prior probability of the hypothesis with the following factor, which is a ratio of how likely an event is given the hypothesis and the prior probability of the event itself:

$$\frac{P(E|H)}{P(E)} \tag{4}$$

To calculate *P(E)*, we use the law of total probability and partition it into disjunct parts:

$$P(E) = \sum_{i=1}^{n} P(E|H_i) \cdot P(H_i) \tag{5}$$

In the application of (1) and (3), it is standardly assumed that we need a prior probability distribution over the set of *all* possible messages, or over the set of *all* possible hypotheses. That is, it is tacitly being assumed that the agents are *theoretically omniscient* as Martins (2005) says—or *modally omniscient* as I like to call it.

However, here I think that a case can be made that this is exactly not how we should think about these equations. Before stating my argument, we note that there has been some precedence for this skeptical attitude in the literature. In itself this does not provide much evidence for the controversial claims that I am going to advance. But when one is fighting against accepted dogmas, it is always comforting to know that one is in good company. In his John Locke lectures, Hartry Field has expressed skepticism about this idea of real agents having a probability function that assigns values to the set of all possibilities.[24] As a result, his favored approach to the problem of logical omniscience (to which we will return to in chapter V) is only to work with probabilistic constraints on rational beliefs (H. Field,

2009). Moreover, in a review of Gallistel & King (2010), Donahoe (2010: 84-85) makes the following interesting observation:

> the relation between information theory and Bayes' theorem was included in a graduate course in experimental design that I taught in the 1960s. However, it became apparent that although this formulation might be useful for guiding the behavior of theorists (cf. Chamberlin, 1880; Platt, 1964), it was inadequate for guiding the ongoing behavior of organisms. When a state is encountered for the first time, neither its prior existence (by definition) nor its a priori probability could be known. The second deficiency may be tolerated because it can be shown that, over repeated occurrences in at least simpler cases, the probability estimate of a state will converge to the population value regardless of its initial value (LaPlace's principle of insufficient reason). The first deficiency is fatal because it presupposes that the organism has foreknowledge of all the states that could ever occur. (…) In communication theory, the possible states are known to the theorist (e.g., the letters of the alphabet) but all the events that an organism may ever encounter in its lifetime are unknowable.

In this passage, Donahoe has already emphasized the crucial point that will occupy us below; to wit, that by using probability functions that are defined on an algebra over the set of *all* possibilities, it is implausibly being presupposed that the system has foreknowledge of all the states that could ever occur, and that it has already determined which probability they should be attributed. Incidentally, he thinks that this assumption is more palatable in the case of the theorist than in the case of biological organisms. But here I am less optimistic, because even if the theorist has a finite set of elements that he needs to know (such as the letters of the alphabet), generating the infinite combinations that it can give rise to, and assigning a prior probability to each combination, still poses an additional challenge.

So to start with equation (1), if we want to apply Shannon's theory of information to modeling communication in biological organisms, or cognitive systems, then we are faced with the problem that it is simply not plausible that they have assigned a prior probability to *any* possible message—no matter how biologically or cognitively *irrelevant* it would be relative to the environment, where the system does its computations. Furthermore, when

turning to equation (3), we are facing the corresponding problem that if we want to model how scientists update their degrees of beliefs in scientific hypotheses (or how individual thinkers update their degrees of beliefs in everyday hypotheses), it is simply not plausible that they should have assigned a prior probability to every possible hypothesis and type of event—no matter how scientifically or cognitively irrelevant it should be for the current theoretical discourse (or for the thinker's background beliefs about the domain of knowledge).

Moreover, theoretical discourses and background beliefs are dynamic entities, which should be treated as temporally indexed and as subject to psychological constraints on the imagination of the knower(s), which prevent them from covering the vast set of all metaphysical possibilities. It is not uncommon for scientists to marvel at puzzling discoveries that they would previously have thought impossible. The course of science is as much a history of *revisions of modal intuitions* about alleged impossibilities and discoveries of new possibilities as it is a history of adjusting degrees of beliefs in hypotheses to the evidence (cf. Nozick, 2001, Stanford, 2013: section 3.3). To take just the most obvious example: if a time travelling machine is ever constructed, skeptics of this point should try travelling back to the nineteenth century and inquire the leading scientists about either the theory of relativity or quantum mechanics. Now as it happens, they might just react with assigning a probability just above zero to all of its assertions. But it is very much an open empirical question, whether they did so before we pose the question. And I am sure that they didn't venture to think that the leading scientists would spend a significant part of the twentieth century on trying to unify such peculiar frameworks with Newtonian physics, and why should they? At the time there was no theoretical discourse, where such a question even made sense. So why assume that the scientific community already had a hypothesis space, where it assigned a prior probability to such weird possibilities.

Moreover, if such questions about revisions of (epistemic) modal intuitions can arise in relation to using equation (3) to model the theoretical discourse of a scientific community, then surely the individual knower will also not have a prior hypothesis space, where different hypotheses about the unification of the theory of relativity and quantum

mechanics with Newtonian physics are assigned a prior probability (regardless of the *Bildung* of the knower).

The alternative is to think about the scientific discourse, and the individual knower, as using a hypothesis space of *serious possibilities (SW)*,[25] which is a subset of the set of all possibilities (*W*), and then to hold that only this subset is assigned probability values (for the purpose of serious inquiry) before the following question arises: if $w \in W$, ought w also be a member of *SW*? That is, the idea is that the inquiry, and the theoretical discourse, only takes place in a subset of the set of all possibilities, which is considered the candidates that are motivated by the present state of knowledge. Only of these candidates does it hold that it would pay off to invest the cognitive and material resources to gather evidence and scrutinize all the implications that the current state of knowledge would have for the possibilities in question.

In the nineteenth century, this time had not yet arrived for the theory of relativity and quantum mechanics. This is why the question of a possible unification with classical Newtonian physics would in all likelihood have appeared ludicrous given the then prevailing state of knowledge and theoretical understanding. At that time, w was not yet a member of *SW*. So it was not yet a possibility that should be assigned a probability (no matter how small) for the purposes of conducting serious inquiry. Rather it was a possibility that deserved to be *ignored* due to its irrelevance for the then dominating theoretical discourse. Subsequent evidence, ability to solve longstanding, theoretical puzzles, and conceptual work on the fundamental categories by original thinkers such as Einstein and Bohr overthrew $SW_{19\text{th c.}}$ and eventually replaced it by $SW_{21\text{th c.}}$

This is the theoretical motivation for thinking of the set of all possibilities over which probability functions are defined, when conducting inquiry, as a subset of the set of all possibilities (*W*), which can decrement and increment.

Here I am not going to attempt to give a complete formal account, however. Rather my concern is primarily with sketching the general idea, and some general ideas about how to implement it. Then it must be up to the formal system-builders to work out whether this implementation works, or whether the underlying thoughts can be better captured in another way.

One formal limitation is that we still need a sample space of mutually exclusive and jointly exhaustive possibilities, where the axioms of the probability calculus are satisfied (i.e. where all the probabilities add up to one etc.). But how can this be guaranteed, if we are working with a subset of W, which in itself does not make up a full partition of W?

One clue was already given in the discussion above; to wit, that for possibilities that are not a member of SW at time t_1, it holds that they simply get ignored until the state of theorizing is such that the prevailing theoretical framework is subject to substantial revisions. One way of implementing this idea would be to hold that for the knower(s) at t_1, the set of all possibilities that are members of SW are simply treated as the set of all possibilities *for the purpose of serious inquiry*.[26] That is, when it comes to assigning prior probabilities, and updating the probability of the hypotheses as a function of the evidence, the set of serious possibilities could be treated as the only possibilities that are taken into account by simply replacing W with SW until further notice. Revision of (epistemic) modal intuitions would then be analyzed as follows: at t_1 the knower(s) assigned a probability of one to the proposition that the actual world was a member of SW_{t1}. At t_2 the set SW_{t1} has been replaced by SW_{t2}, and now the knower(s) no longer assign a probability of one to the union of all the elements in $SW_{t1.}$

Another apparent obstacle to making good of the idea of replacing W with SW is this: that something with a probability of zero would have to occur, when the knower(s)' (epistemic) modal intuitions are corrected. Before the knower(s) took it that the probability of the union of all the propositions defined over SW_{t1} was equal to one. Now an event has occurred (or a proposition is assigned the value 'true'), which was before assigned a probability of zero (because the proposition A isn't a member of the algebra, S, that is defined over SW_{t1}) *for the purposes of serious inquiry*. So how are we to think about its occurrence?

If P is a probability measure that is defined on $\mathcal{A}$, where $\mathcal{A}$ is an algebra over W, then we can consider SP a probability measure on a finite subalgebra $S \subseteq \mathcal{A}$, where it holds that the values assigned by P to members of $\mathcal{A}$ represent the knower(s) subjective degrees of belief taken in the abstract.[27] In contrast, the values assigned by SP to the members of S

represent the degrees of belief that the knower(s) assign to the possibilities under consideration for the purposes of serious inquiry.

When the agent is not conducting serious inquiry, he has a different probability distribution function. For instance, in his own practical reasoning, he might assign a probability of one to the belief that the stock market crashes every x^{th} year, or that the air plane that he is about to enter won't fall down (because he treats these propositions as *suppositions* in his own practical reasoning). In effect, these propositions will then count as *practical necessities* for him in those contexts (i.e. they are treated as having a probability of one in spite of the fact that he knows very well that he doesn't possess the corresponding evidence, which would allow him to justify assigning such a high probability to others). Or another way of expressing this idea would be to say that the agent is *existentially committed* to these propositions to the point, where he is willing to risk the financial prosperity of his whole family on his beliefs about the stock market, and his own life on his belief in the safety of flying with that particular air plane. However, if this agent is a statistician, then he knows better than to write academic papers, where he assigns the corresponding probabilities to the propositions in question, because whatever evidence he may have wouldn't hold up to the academic standards of his discipline in spite of the fact that they may ground his subjective certainty.

So to return to our question: is it a problem that the proposition, A, isn't a member of the algebra, S, that is defined over SW_{t1}, and that it is consequently assigned a probability of zero before t_2, when it becomes a member of SW_{t2} (again: for the purposes of serious inquiry)? It would be a problem if the assignment of a probability of zero was thought of as representing *metaphysical impossibility*, because then something that was supposed not to be capable of existing in any possible world suddenly popped into existence; a contradiction in terms. But it is not problematic, if assignment of a probability of zero is thought of as representing *epistemic impossibility*. Indeed, it happens all the time that things happen which knower(s) take to be impossible, because they have faulty (epistemic) modal intuitions. So in relation to the issue above, there should be no problem with making sense of the notion that there is a state of knowledge at t_1, and a state of play in the theoretical discourse at t_1, where the truth of A is treated as an (epistemic) impossibility for the purposes of serious

inquiry, but where this assumption of the discourse is later retracted due to some new developments.

Moreover, in general there is no room for skepticism about events occurring that have a probability of zero, or of events not occurring with a probability of one. For in a continuous probability space each individual point is assigned a probability of zero. Yet, we know that one of them will be the outcome, and probabilistic coherence requires that we assign a probability of one to the occurrence of one of the remaining points. Yet, none of them will occur, if this point is the actual outcome.

So for purposes of general philosophical theorizing, one suggestion is that we adopt the following picture: we can think of there being a sampling space of *metaphysical possibilities*, *MW*, which is so vast that no community of epistemic agents is able to investigate more than a small corner of it. And then we can think of there being a range of subsets of sampling spaces of *MW* (e.g. SW_{t1}), which represent the *epistemic possibilities* that the agents take into consideration in their inquiries (and again the point is that for different purposes of inquiry, they can use different sampling spaces).

To illustrate the idea with an example borrowed from Eric Raidl (personal communication): even when we consider a simple experiment like throwing a coin and assigning probabilities to the different possible outcomes, we never include *all* the logical possibilities (in spite of our occasional rhetoric to the contrary). Rather, we just consider events like the coin landing up heads or tails as forming a partition of the space of all possibilities in spite of the fact that we know very well that it would not count as a partition of the space of metaphysical possibilities. For there are always the extremely rare and outrageous possibilities like the coin landing on the edge, or it being caught up in some weird magnetic field and never landing again, which are never taken into consideration. So even the sampling space, *W*, is nothing but a small corner of *MW* (in spite of the fact that *W* is normally stipulated to be exhaustive).

The challenge for the philosopher of science, the sociologist of science, and the psychologist, would then be to identify the conditions under which the hypothesis space of serious possibilities is replaced. The underlying idea is that not every logically possible

hypothesis is interesting enough for us to update our beliefs in it on the basis of the evidence. As the physicist, David Deutsch, says:

> Science would be impossible if it were not for the fact that the overwhelming majority of false theories can be rejected out of hand without any experiment, simply for being bad explanations (2011: 25).

There are in other words implicit standards in science for when a hypothesis has been established to be interesting enough for it merit to be investigated.[28]

The challenge for the formal epistemologist is to find some way of explicating how the hypothesis space of serious possibilities is replaced, which doesn't rely on simple conditionalization, because if the probability 0 is assigned to any possible hypothesis, then it can never rise again and gain a positive probability by means of simple conditionalization. As of this moment, my main concern is with establishing the methodological points expressed in this chapter. So I am happy to delegate such challenges to colleagues in the hope that this appendix has at least succeeded in articulating some new, interesting ideas that deserve further exploration.

---

[2] Acknowledgements: this chapter profited greatly from discussions with Eric Raidl, Edouard Machery, Sieghard Beller, Wolfgang Spohn, Björn Meder, Thomas Müller, and the members of two colloquia at the University of Konstanz.

[3] Examples of controversial operationalizations: one particular striking example is the philosophical reactions to the experiments of Benjamin Libet, where part of the complaint was that it is unclear that liberty as it is involved in making the most important decisions of our lives can be fully operationalized in terms of the onset of simple decisions about the movements of body limbs (cf. Sturma, 2006). Another example is the operationalization of executive functioning through tests that merely measure the participants' ability to follow externally supplied rules while ignoring distracting stimuli, which Stanovich (2011: 56-59) argues doesn't really deserve to be designated as 'executive functioning'.

[4] Exception to lower power of non-parametric tests: if the sampling distribution is not normally distributed then non-parametric tests need not have a lower power than parametric tests (A. Field, 2009: 551, Howell, 1997: 646).

[5] On the methodology of pure philosophy: in fact, in Olsen (2014), I take a first stab at articulating a methodology for pure philosophy. So I am not hostile to such projects.

[6] Reference: two representative examples of this more traditional contribution of philosophy that deal specifically with psychology of reasoning are Stein (1996) and Samuels, Stich, & Bishop (2002).

[7] Reference to reflections on actual scientific practice: Platt (1964) provides a valuable resource for how this method of exclusion of possibilities takes place from the perspective of a working scientist.

[8] Quote: Platt (1964: 351) says in relation to the empirical discourse that: "A failure to agree for 30 years is public advertisement of a failure to disprove". Of course, such a timetable does not put the philosophical discourse in a particularly favorable light, where some discussions have been going on for over two millennia.

[9] Justification: it is easily shown that surprising events (which have a low prior) convey the most information in Shannon's theory of information and enables the largest updates in the hypotheses that predict them according to Bayes' theorem (cf. Gallistel & King, 2010: ch. 1-2). If in turn the occurrence of an event becomes less surprising the more plausible theories we have that are capable of predicting it, it holds that predictions that are unique to a theory will provide stronger confirmation for a theory than if they were shared by other theories due to their lower prior.

[10] Explication on the axis-tilt theory of seasonal change: briefly stated, this theory holds that the seasonal change is to be explained by the fact that the earth's axis is tilted around 23 degrees relative to the plane in which the earth orbits around the sun. As a result of this tilt, there is an asymmetry in when the two hemispheres are closer to the sun as well as in the angle by which different parts of the earth are reached by the rays of the sun at any particular time, which explains why seasons change.

[11] Further examples of philosophical theories that have been applied in science: further good examples include: (1) the influence Wittgenstein's considerations on family resemblance had on the prototype theory, (2) Carrey's (2011) development of Quinian bootstrapping as a model for cognitive development, (3) Dennett's suggestions about how to test whether chimpanzees have a theory of mind, which was implemented in the famous false belief task by the psychologists Wimmer and Perner (Andrew, 2012: 21), (4) the influence of Fodor's theory of modularity on cognitive science (Nichols & Stich, 2003: 117), (5) the implementation of Thagard's work on explanatory coherence in the computer model ECHO (Markman, 1999: 111ff.), and (6) the application of the Ramsey test, the equation of P(if p then q) with $P(q|p)$ in Evans & Over (2004) and Oaksford & Chater (2007) as well as the latter's use of Bayesian confirmation theory. But interestingly, Frixione and Lieto (forthcoming) argue that a central case where this adoption of philosophical ideas for scientific purposes has not happened successfully is in regard to the distinction between conceptual/nonconceptual due to the differences in concerns between philosophers and psychologists in relation to concepts.

[12] Reference: Murphy (2004: ch. 2).

[13] "Implication 4 as a possible exception": implication 4) is a bit of an exception in this context; however, failures of transitivity have also been reported, as Murphy (2004: ch. 2) points out.

[14] Non-monotonic reasoning and the typicality effect: although Murphy doesn't discuss it, it is probably reasonable to assume that the much discussed example about Tweety the bird in the literature on non-monotonic reasoning (cf. Frankish, 2005) is simply due to the typicality effect. Once the concept 'bird' is used, the information from long-term memory most readily accessed consists of information about typical birds that fly, which is why the default assumption is that Tweety flies upon learning that it is bird. In comparison, penguins are atypical members of this category, so although we all know that they would provide a counterexample to the inference that we are prepared to make, the conceptual information needed for realizing this much is not accessed as readily. [As it turns out, I later discovered that Gärdenfors (2005) had seen the same connection before I did.]

[15] Category-based induction: in an experimental design originating with Lance J. Rips, it is a robust finding that subjects are more inclined to infer that other birds will become infected by a disease upon learning that a typical member of the category (e.g. robin) has it than when learning that an atypical member (e.g. eagle) has it (Murphy, 2004: ch. 8, Machery, 2009: ch. 7).

[16] On summary representations: being a *summary* representation means that the entire category is represented by a unifying representation (in the sense of a description of the category as a whole), and this representation is *weighted* to allow for some properties to be more important than others—even if none of them are themselves necessary or sufficient for category membership. It should moreover be noticed that the prototype theory has been developed in many versions, which, *inter alia*, differ in terms of the character of the statistical knowledge that goes into computing the weight of the properties represented in the summary representation (see Murphy 2004: ch. 2-3, p. 109ff., Machery 2009: ch. 4). However, for a first pass it may suffice to think of a simple version that merely uses information about the relative frequency with which members of the category have particular properties.

[17] Reference: Rey (1985, 2005).

[18] Responding to a potential objection: to be sure, Murphy (2004: ch. 2) does consider a parallel line of argument, but then argues that it probably only applies to small, closed, and person-made systems like chess, because in all the examples he considers, vagueness makes its ugly appearance whenever one scratches the surface. However, considering all the examples Smith mentions, I think that it is fair to say that Murphy's argument is at best suggestive at this point, and I therefore choose to let the point stand.

[19] Reference: see also Carruthers' (draft) criticism of the failure of domain-general theories of rationality in philosophy to pay heed to the requirements of working memory.

[20] Qualification concerning computational models: it should be noted that the example of Oaksford & Chater (2007) shows that it is possible for a psychological theory to generate predictions without specifying underlying cognitive processes. But at least then they provide a formal, computational model, which is not something that has been on the agenda in philosophy.

[21] Reference: see, for instance, Oaksford & Chater (2007: 14), Johnson-Laird (2008: ch. 4-5), and Gallistel & King (2010: 32).

[22] Reference: see Tedeschi & Felson (1994: ch. 1,2,6).

[23] Reference: for details see Gallistel & King (2010: ch. 1).

[24] Reference: http://podcasts.ox.ac.uk/people/hartry-field

[25] On the origin of the notion of a serious possibility: the notion of a serious possibility is one that is to be found in Levi (1991, 1997) (see also Spohn, 2006). But independently of Levi, I found it useful to think about inquiry as taking place in a space of hypotheses that have been established to deserving the status of demanding to be taken seriously in Olsen (forthcoming).

[26] Examples of other types of discourse: for the purpose of amusement (e.g. writing fiction or making jokes), the members of *SW* need not be treated as the set of all possibilities.

[27] Note on an problematic assumption in contemporary work: actually, I am not even sure how much sense this notion of having a measure of degrees of belief in the abstract really makes, where we don't take the cognitive and practical context into account. But since it is these terms in which formal epistemology are usually conducted, I will do so as well here for expositionary purposes.

[28] A useful case study: as I have suggested in Olsen (forthcoming), a good case study for pursuing this topic is Smolin's (2008) attempt of establishing alternatives to string theory as candidates that deserve to be taken seriously.

II

Motivating the Relevance Approach to Conditionals[29]

Abstract: In chapter one, some very general considerations were introduced about how to make philosophical theories about cognitive competences useful for the empirical sciences. In the next three chapters, we shift gears by implementing these recommendations with respect to the ranking-theoretic approach to conditionals. The first step in this process is to theoretically motivate the relevance approach to conditionals on which the latter is based through a comparative discussion with the other main contenders in contemporary psychology of reasoning. In chapter three, we will have ample opportunity to study the formal details of offering a relevance approach, and how to formulate a mathematical model on the basis of it, which can be of use for experimental psychology. But first we start by exploring some of the key theoretical ideas in the absence of more detailed empirical considerations, and by arguing why a relevance approach is theoretically better motivated than the suppositional theory. In the course of this discussion, an argument will be presented of why failures of the epistemic relevance of the antecedent for the consequent should be counted as a genuine semantic defect (as opposed to be relegated to pragmatics). Furthermore, strategies for dealing with compositionality and the perceived objective purport of indicative conditionals will be put forward.

1. Introduction

Due to accumulating psychological evidence of poor logical performance (Evans, 2002, Manktelow, 2012), and difficulties in making sense from an evolutionary perspective of an ability to deal with necessities as opposed to uncertainties (Oaksford & Chater, 2007), there has been a paradigm shift in the psychology of reasoning (Evans, 2012). Whereas earlier research used deductive logic as the main normative model, recent research has started to use probabilistic, Bayesian models.

In the study of conditionals this is seen by a shift away from approaches that analyze the natural language conditional in terms of the material implication ('$\supset$') towards probabilistic models that represent our understanding of the natural language conditional as a conditional probability (Evans & Over, 2004, Oaksford & Chater, 2007: ch. 5, Oaksford & Chater, 2010a). In the following, we shall cover some of the theoretical background for making this move.

1.1 The Horseshoe Analysis

The traditionally held view that the semantic content of the natural language indicative conditional is determined by the truth-conditions specified by the table below has somewhat disparagingly been called the 'horseshoe analysis':

A	C	$A \supset C$
$\top$	$\top$	$\top$
$\top$	$\bot$	$\bot$
$\bot$	$\top$	$\top$
$\bot$	$\bot$	$\top$

One of the theoretical difficulties with accepting this account is that it forces us to analyze conditionals with false antecedents as true no matter what the consequent. As a result, both of the following conditionals are to be treated as true in spite of intuitively felt problems with treating (2) as true and the fact that treating both as true would violate a requirement of conditional consistency:

If Sahara is covered with ice, it is cold in Sahara. (1)

If Sahara is covered with ice, it is warm in Sahara. (2)

A further difficulty is that the abovementioned analysis validates a number of argument schemes that are hard to accept. One case in particular is the so-called paradoxes of the material implication. Since the material implication is guaranteed to be true, whenever either the antecedent is false or the consequent is true, the following argument schemes are valid according to the horseshoe analysis:

$$\frac{\sim A}{\therefore \text{if } A, \text{C}} \qquad \frac{C}{\therefore \text{if } A, \text{C}} \qquad [1]$$

This gives rise to paradoxical outcomes, when natural language content is substituted into [1] as it would permit the formulation of arguments with any arbitrary degree of absurdity such as 'it is not the case that Anders Fogh is the prime minister of Denmark. Hence, if Anders Fogh is the prime minister of Denmark, then Konstanz has direct access to Bodensee' and 'Konstanz has direct access to Bodensee. Hence, if Anders Fogh is the prime minister of Denmark, then Konstanz has direct access to Bodensee'.[30]

But in rejecting the validity of [1], one has to be aware that one is thereby also committed to rejecting the universal validity of the *or-to-if inference* as its validity would entail the validity of [1]:[31]

$$\frac{A \vee C}{\therefore \text{if} \sim A, \text{C}} \qquad [2]$$

However, although this inference may be reasonable in contexts, where one has evidence for either A or C without knowing which one holds, and one is in the process of eliminating alternatives, it cannot be accepted as a general principle of reasoning (cf. Spohn, 2013a). The reason is that in combination with disjunction-introduction, it would validate the following argument schema:

$$\frac{\begin{array}{c} A \\ A \vee C \end{array}}{\therefore \text{ if } \sim A, \mathrm{C}} \qquad [3]$$

Accordingly, if one is in a position to treat A as true, then one would have to accept the conclusion that if A were not true, then C—*for any arbitrary proposition C*. But then we are back at the absurdity that [1] left us with.

In addition to [1] and [2], the horseshoe analysis also validates a range of further argument schemes with conditionals in the conclusion like strengthening of the antecedent (if A, C ∴ if A & B, C), contraposition (if A, C ∴ if ~C, ~A), and transitivity (if A, B; if B, C ∴ if A, C), which have counterexamples that are well-known in the literature (cf. Bennett, 2003: ch. 9).

1.2 The Suppositional Theory of Conditionals

In rejecting the horseshoe analysis and favoring a probabilistic model, the experimental literature is following a philosophical tradition that has, *inter alia*, found expression in the works of Adams (1965), Edgington (1995, 2003, 2006), Woods (1997), and Bennett (2003). Although there are individual differences, these theoreticians are unified by a commitment to the *suppositional theory of (indicative) conditionals*, which consists of the following core claims:

> *The Ramsey Test*: conditionals are assessed by temporarily adding the antecedent to one's knowledge base and evaluating the consequent on the supposition that the antecedent is true.
>
> *Adam's Thesis*: P(if A, C) = P(C | A), where 'A' and 'C' are not allowed to be conditionals in turn. Or rather: Acc(if A, C) = P(C | A).
>
> *P-validity*: the validity of arguments containing conditionals in the conclusion is not to be assessed on the basis of deductive validity but on the basis of p-validity. The inference from the premises, $A_1, A_2, \ldots, A_i$, to the conclusion, *B*, is *p-valid* just in case the uncertainty of the conclusion cannot exceed the sum of the uncertainty of the premises: $[1\text{-P}(A_1)] + [1\text{-P}(A_2)] + \ldots + [1\text{-P}(A_i)] \geq 1 - \text{P}(B)$.

One issue deserves comment. The reason why P(if A, C) = P(C|A) is replaced by Acc(if A, C) = P(C|A) in Adam's thesis is due to Lewis' triviality results, which showed that in general no proposition can be found, whose probability is equal to P(C|A) for all probability distributions, without their being subject to trivializing features such as only being able to assign positive probability to two pairwise incompatible propositions and collapsing P(C|A) to P(C) (Bennett, 2003: ch. 5, Woods, 1997: ch. 4, p. 114-8). However, if indicative conditionals don't express propositions, then they can hardly be assigned probabilities in the literal sense, and it is thus arguable that Adam's thesis should be formulated in terms of the *assertibility* or *acceptability* of a conditional, instead of in terms of its probability. Hence, strictly speaking P(if A, C) = P(C|A) should be replaced by Acc(if A, C) = P(C|A) in formulating the suppositional theory of conditionals (cf. Douven, forthcoming: ch. 3).[32] But in practice this subtlety has not been observed in the psychological literature.

However, there are some differences over, whether Lewis' triviality results should make us reject the idea that conditionals have truth values, or whether it is possible to maintain a deflationary sense in which conditionals have truth values according to the de-Finetti truth table:[33]

A	*C*	*If A, C*
⊤	⊤	⊤
⊤	⊥	⊥
⊥	⊤	*void*
⊥	⊥	*void*

If the latter position is adopted, supposedly a three-valued logic would have to be formulated. However, the proponents of this variant of the theory have not in general carried out such a program.

One attractive feature of the suppositional theory of conditionals is that the notion of p-validity allows us to reject the argument schemes that were found problematic in section 1.1. In contrast, arguments not containing conditionals in the conclusion will be valid according to p-validity, if they are classically valid.

As Bennett (ibid: 139) says in relation to the paradoxes of the material implication, the present theory does:

> imply that when my value for P(A)=0, I should have no value for P(if A, C); and that when my value for P(C)=1, my value for P(if A, C) should be 1. But nothing follows about the value for P(if A, C) when P(A) is low but > 0, or when P(C) is high but < 1. [Notation changed to yield uniformity, NSO.]

However, a case could be made that it doesn't invalidate [1] for the right reason. The most natural analysis of its defect is that the problem lies in: (a) validating inferences to conditionals that seem unsupported by the premises, and (b) allowing the antecedent to be *irrelevant* for the consequent in the conditionals introduced.

So although it preserves truth from the premises to the conclusion to make inferences like those involving Anders Fogh from section 1.1, making them would violate our expectations about the epistemic relevance of the antecedent for the consequent to such a degree that it might be doubted that the speaker really understood what he was saying.

Viewed from this light, the ability of the suppositional theory to invalidate [1] by pointing out, as Bennett does, that no constraints on $P(C|A)$ are made, when P(A) is low but > 0 and P(C) is high but < 1, is a small victory as it likewise fails to accommodate (a) and (b). Furthermore, as the same quote reveals, the suppositional theory is actually committed to the permissibility of the inference to 'if A, C' from 'C', whenever P(C) = 1. But, of course, then the door is once more open to introduce absurd inferences, where the antecedent is irrelevant for the consequent. Moreover, if the conclusion consists of conditionals containing candidates for "analytical" connections between the antecedent and the consequent, p-validity permits that they be inferred from any premise where $P(A) \leq 1$. Hence, the inference to 'if England has a queen, then the royal family in England has at least one female member' from any arbitrary premise (no matter how irrelevant) is sanctioned. So not only does the suppositional theory of conditionals fail to render [1] invalid due to (a) and (b), but it would appear to have its own problems with satisfying these constraints as well.

A further related worry is that the de Finetti truth-table seems to be a small improvement, when it comes to handling our pair of Sahara conditionals from section 1.1. The reason is that holding that (1) and (2) take the truth value 'void' due to their false antecedents leaves us unable to explain the defect of (2) with its insinuation that Sahara's being covered by ice somehow constituted a reason for thinking that Sahara is a warm place to be.

It should finally be noted that although I didn't include these further claims among the core theses above, often the suppositional theory is stated by saying that:

(i) assertive uses of conditionals are to be understood as *conditional assertions*, where one is only prepared to assert the consequent under the supposition of the antecedent being true (Edgington, 1995, Woods, 1997, Bennett, 2003: ch. 8, and Oaksford & Chater, 2007: ch. 5),

(ii) the suppositional theory captures Ryle's (1950) idea of conditionals serving as inference tickets (Bennett, 2003: 118, Oaksford & Chater, 2007: ch. 5), and

(iii) a conditional is rationally acceptable iff the conditional probability is above some threshold.

However, in appealing to intuitively attractive ideas like (i) and (ii), the proponents of the suppositional theory are invoking auxiliary hypotheses that are not systematically related to what has been identified as the core claims of the theory above. For it would appear that a more natural home for (i) and (ii) would be an account emphasizing that the antecedent must be (epistemically) relevant for the consequent, and that the antecedent may favorably be thought of as a reason for the consequent in assertive uses of conditionals, which is the kind of approach we turn to in the next section.[34]

Finally, a problem with (iii) is that the conditional probability, P(C|A), will be high whenever P(C) is high and the two propositions are probabilistically independent of each other. Hence, (iii) can be satisfied even when there is no epistemic connection between A and C. So if the argument in section 2 is successful that failures of epistemic relevance

should be considered semantic defects of conditionals, then (iii) will not do (see also Douven, forthcoming: ch. 4).

1.3 The Relevance Approach

There is an older tradition in philosophy to understand the paradigmatic cases of natural language conditionals as expressing inferences, where the premise is a reason for the consequent, which in recent times has been articulated by Goodman (1991, [1947]), Ryle (1950), Rott (1986), Strawson (1986), Brandom (1994), and Douven (2008, 2013, forthcoming) and made precise using the formal tools of ranking theory in Spohn (2013a, forthcoming). Until now this view has found little play in psychology of reasoning, and one of the main purposes of the chapters to follow is to prepare the way for such application.

Viewed from the present perspective, p-validity shares the same problem with deductive validity in that it defines validity in terms of a formal property that permits us to draw inferences to conclusions that don't preserve reason relations from the premises to the conclusion. Historically, the realization that the premises must somehow be relevant for the conclusion was first captured by Alan Ross Anderson and Nuel Belnap in relevance logic by the syntactical constraint that the premises are actually used in the construction of a proof for the conclusion (cf. Mares, 2007: 6ff.).

Relevance logic also invalidates [1] and [2]. However, the psychological literature cited in the beginning of this chapter makes it doubtful how fruitful this strategy of modeling everyday reasoning on the basis of an analogy with mathematical proof really is (cf. ibid: 29). It is thus an attractive feature of the account offered in Spohn (2013a, forthcoming) that some of the intuitive ideas originally motivating relevance logic can receive a new life with a probabilistic understanding of relevance (and reasons) as consisting in probability difference making (i.e. $P(C|A) \neq P(C|\overline{A})$, where $\overline{A}$ expresses the negation of the proposition, A).[35]

A symptom of the difference between the two approaches is that whereas Mares (2007: 44) attempts to ground situated inferences on informational links that have to be *perfectly reliable* (to the exclusion of most causal relations), the relation of probability raising allows the cognitive system to put a weight on how strong the association in the

information links is and to exploit links that are only reliable enough to be useful. The latter approach is thus in a better position to connect with the general orientation towards reasoning under uncertainty with degrees of beliefs currently dominating psychology of reasoning than the former.

However, as the two approaches are based on the same intuitions, we can follow Mares (2007: 14) in citing the following examples as a way of motivating why the dimension of relevance should be integrated into our semantic analysis. Knowing that guinea pigs have no tails, we would probably find that there is some sort of semantic defect in the following indicative conditional:[36]

If I pick this guinea pig up by the tail, its eyes will fall out. (3)

Since we know that the antecedent is false, it seems problematic for the conditional to suggest that there is some sort of connection between guinea pigs being picked up by their (non-existing) tail and their eyes falling out.

Of course, opponents of the relevance approach would hold that it is just pragmatically misleadingly to use this conditional in certain contexts, because it carries this implicature. However, in reply we must ask, why the opponent of the relevance approach is so certain that there is a layer of semantic competence in ordinary people, whereby a (literal) meaning can be attributed to conditionals of this type, without it appearing that there is some sort of semantic defect. This is surely an empirical question that cannot merely be decided by the intuitions of card-carrying theoreticians.

Moreover, as Mares (ibid.) points out, the indicative above also has a counterfactual analogue. So the problem cannot just be set aside as a local one without implications for the core theory:

If I were to scare this pregnant guinea pig, its babies would be born without tails. (4)

Again here it seems that this conditional should be treated as having some sort of semantic defect merely in virtue of the fact that the babies of guinea pigs will in any case be born without tails.

In addition, there are the examples of semantic defects due to the antecedent's obvious irrelevance for the consequent, where both are true, which will have to be accepted as having a literal meaning, whereby they are perfectly fine, according to the horseshoe analysis and the suppositional theory (cf. Edgington, 1995: 267):

If Napoleon is dead, Oxford is in England (5)

Before justifying why lack of relevance should be considered a genuine *semantic defect* of conditionals in section 2.2, section 2.1 will provide an argument against attempts of restricting considerations of relevance to a pragmatic component that should not be allowed to enter into our semantic analysis.

2. The Semantics/Pragmatics Distinction

To set the stage, the following quote is instructive as it laments the tendency to ban items from our semantic analysis merely because the dominant semantic theory is unsuitable to handle them:

> *Method determines Matter*: If we are to say what an expression means by giving truth conditions, then "Goodbye." has no meaning. If we are to say what an expression means by describing its use, then "Goodbye." does have a meaning. I believe that the tendency to banish a wide variety of semantic regularities (including those of indexicals) to the netherworld of 'pragmatics' has been a direct consequent of the fact that the dominant forms of semantic theory are unsuitable for these expressions. (Kaplan, unpublished: 4)

Many things are controversial in philosophy of language. One thing in particular is the issue of where to draw the distinction between semantics and pragmatics. However, we are forced to confront this issue as there is a tendency in the literature on conditionals to grant that there is a strong intuitive force in saying that there must be some sort of connection between the antecedent and the consequent (or that the first must be

epistemically relevant for the second), but then to set this issue aside as a topic to be dealt with in pragmatics in stating the semantics of conditionals (e.g. Edgington, 1995: 269).

One tradition focuses on the descriptive use of language as the core of meaning to be accounted for first on which everything else builds (e.g. the account of speech acts and non-literal uses of language), and analyzes it in terms of truth conditions of sentences that are constructed on the basis of referential relations to the world. In this tradition, a strong distinction between semantics and pragmatics can be made by holding that truth conditions give the context-invariant meaning of sentences, which can then be modulated pragmatically through factors holding in particular contexts. Yet, the picture gets complicated by the fact that even things like reference assignment and scope interpretation, which are needed for specifying the requisite truth-conditions, may depend on pragmatic considerations (Riemer, 2010: 129).

Another tradition holds that meaning is to be understood in terms of use and that we should understand the meaning of sentences in terms of an analysis of the pragmatics of their appropriate use (cf. Brandom, 1994, Khlentzos, 2004). Within this tradition, it will still be possible to draw a distinction between effects of *mere pragmatics*, and the general analysis of meaning, by holding that: (a) the pragmatics of appropriate use can be modulated by context-specific factors, and (b) there are norms of appropriate use like norms of politeness or prudence, which are not to be included in the analysis of meaning as it only focuses on epistemic norms. (And again the idea is that this assertive use of language is the core of meaning that is to be accounted for first on which everything else builds.)

In both cases, it then seems attractive to set aside effects of mere pragmatics in the general analysis of meaning, which is supposed to deal with contents that can be assigned to sentences on the basis of the linguistically (e.g. syntactically) encoded information and only a bare minimum of knowledge about context-specific factors. We can still include indexicals (e.g. 'I', 'here') in the semantic analysis on the basis of this criterion by holding that the kind of pragmatic modulation referred to above goes beyond the setting of values of semantic variables by changing the standard meaning that can be assigned to a phrase/word/sentence through extra premises that only hold in a given context.

2.1 Reason Relations as Part of the Sense Dimension of Meaning

While it remains controversial in philosophy of language, whether both items in Frege's distinction between *sense* and *reference* are to be included in our theory of meaning, it is common in linguistics to adopt this distinction and to think of the latter as dealing with the relation between language and the world and the former as dealing with a relationship between elements within the vocabulary system (Saeed, 2003: section 1.6.1). In the case of word meaning, it is standard to take *sense* as encompassing lexical relations like synonymy, antonymy, and meronymy (i.e. whole-part relation) (ibid: ch. 2-3).[37]

Of course, in this the linguists are going beyond Frege, who mainly dealt with sense in relation to informative identity statements. But we can take him as having discovered the general need to include the cognitive role that linguistic content plays in information processing as part of the semantic analysis and then go beyond him in specifying what parts of linguistic content contribute to its cognitive utility. Furthermore, we will also go beyond Frege in pointing to aspects of the sense dimension of meaning, which don't play a role in determining reference.

One suggestion would be to include reason relations under the sense dimension of meaning. There are at least two reasons for making this move. One is that it is possible to consider synonymy and antonymy as themselves involving reason relations. So if 'autumn' and 'fall' are *synonymous*, then the proposition that it is autumn is a reason of maximal degree for the proposition that it is fall. And if 'large' is an *antonym* of 'small', then the proposition that x is large is a reason of maximal degree against the proposition that x is small. Another consideration in favor of including reason relations among our sense dimension of meaning is that there is a list of utterance modifiers like the following, which clearly have a meaning in terms of commenting on the dialectical role of assertions, which is not captured by a truth-functional analysis:

> 'after all', 'besides', 'be that as it may', 'furthermore', 'however', 'in conclusion', 'indeed', 'moreover', 'at any rate',[38] 'still', 'although', 'yet', 'the reason is that', 'on the one hand…, on the other,…', 'thus', 'hence', 'in fact', 'to be sure', 'so', 'consequently', 'in spite of the fact that', 'despite', 'since', 'due to the fact that', 'provided that', 'as a result', 'on the contrary', 'in contrast', 'accordingly', 'whereas' and 'nevertheless' etc.

To introduce a term for building up a dialectical structure by means of these utterance modifiers, I would suggest that we talk about the dialectical compositionality of an argument in addition to the traditional, truth-functional compositionality of a sentence. The expression of reason relations makes up a central element of composing the dialectical structure of an argument, and it seems that it would be part of the linguistic competence of mature language users to be able to decode this dialectical structure on the basis of the modifiers listed above (even when provided with impoverished contextual information).

In the tradition of truth-conditional semantics, it has always appeared attractive to treat 'and' and 'but' alike in that both could be treated as contributing to the composition of sentences as the logical connective 'conjunction'. The caveat is then added that there was a pragmatic distinction between the two consisting in that the latter indicates a contrastive relationship between the conjuncts, which was absent in the former as illustrated by sentences such as 'she was poor but honest'.

However, as we saw above, we already have a wider class of utterance modifiers, whose semantic content in building up a dialectical structure of a text or a conversation cannot be captured in terms of truth-conditional semantics. Hence, it would seem that there is no reason why we shouldn't include this contrastive content of 'but' in its semantic analysis by saying that an expectation about a relationship between being poor and dishonest is being contradicted in the assertion of 'she was poor but honest'.[39] So although the contrastive content expressed by 'but' makes no contribution to the truth-functional compositionality of sentences, it earns its keep in the semantic analysis through its central contribution to the dialectical compositionality of arguments.[40]

A further consideration in favor of thinking that this component should be made part of the semantic analysis of 'but' is that it is a context-invariant feature of its meaning, whose interpretation requires little contextual information. In fact, all we were provided with above was a single sentence and yet its content was fully understood.

Once we have accustomed ourselves to the general idea of including reason relations among the sense relations, the door is open to consider it part of the semantic content of conditional connectives like 'if…then' and 'even if' that they linguistically encode reason relations in a context-invariant manner, which ordinary speakers are capable of interpreting,

when provided with even impoverished contexts. At least, there should be no general resistance to this idea on the grounds that it can only be part of the semantic analysis to ascribe truth-conditions as this leaves us without a semantic analysis of the utterance modifiers mentioned above in any case. Ultimately, it must then be considered an empirical question, whether this semantic analysis is adequate.

2.2 On Semantic Defects

Having thus provided an argument for why we should not be predisposed to reject the idea of expression of reason relations as being part of the semantic analysis in general, we can now return to the explanatory challenge of accounting for why a failure of relevance should be counted as a semantic defect of conditionals.

That card-carrying theoreticians, who are committed to opposing positions, make assertions to the contrary is hardly decisive and we should not let the dispute be left for such theoretically tainted intuitions to decide. So how are we to adjudicate in this theoretical dispute?

One possibility would be to reflect on the nature of semantic defects for the semantics of conditionals. Surely the most obvious sense in which a sentence could have a semantic defect would be, if its truth conditions were such that it could only take the value 'true' or 'false', when in fact the semantic intuitions of ordinary speakers dictates that it should take the opposite value. Another example of a semantic defect would be cases of syntactically well-formed sentences that nevertheless fail to express a proposition that can be evaluated without further pragmatic supplementation. In Bach (1997) and Blackmore (2004: ch. 1-2) cases of this kind are discussed. However, both of these suggestions are only helpful for present purposes to the extent that conditionals in fact have truth conditions, which is itself disputed territory.

If we then set aside problems with truth conditions for the moment, we can consider, whether there are other types of semantic defects that could be used to adjudicate in the dispute at hand?

Since we have already seen that proponents of the suppositional theory have flirted with the idea of conditionals being used as inference tickets, and of the speaker only being

prepared to assert the consequent on the supposition that the antecedent is true, it seems that they cannot object to using the lack of literal meaning that would enable such uses as an indicator of a semantic defect.

Using this criterion, the proponents of the suppositional theory cannot object to treating the examples discussed in section 1.3 as genuine instances of semantic defects as failure to express reason relations blocks the use of conditionals in sensible inferences and the desirability of making the corresponding conditional assertions. Presumably ordinary subjects would fail to identify a sensible commitment to answer justificatory challenges, which could be attributed on the basis of assertions of conditionals, where the antecedent is blatantly irrelevant for the consequent as in (5). And presumably ordinary subjects would fail to take a line of reasoning seriously that involved making use of such conditionals. If so, then the problem with these conditionals is not just that it would violate some Gricean, pragmatic maxim of non-misleading discourse to introduce them in a conversation. Rather their defect consists in having a literal content, which ropes them of their cognitive utility.

So to summarize, the argument for counting epistemic irrelevance as a semantic defect of conditionals has been:

(P_1) We should include sense (cognitive utility) as linguistically encoded in sentences in a context-invariant way, which can be interpreted on the basis of impoverished contexts alone (where little or no supplementing contextual information is provided) as part of the literal meaning of expressions.

(P_2) The ability to express reason relations is a central way of enabling cognitive utility.

(C) Hence, if it holds for conditionals with antecedents that are blatantly irrelevant for the consequent that they don't enable the cognitive utility on the basis of their literal, context-invariant meaning that conditionals expressing reason relations would enable, then the former count as semantically defective.

As it stands, the argument only directly addresses the issue of whether cases of blatant irrelevances like example (5) should be counted as genuine semantic defects. As such, it has been silent on examples (3) and (4). However, it can easily be extended to cover

such cases as they suffer from the defect of stating spurious relationships involving the non-existent tail of guinea pigs, which will prevent such conditionals from having the cognitive utility that normal conditionals enjoy. Accordingly, there is no temptation to using such conditionals as inference tickets and it makes little sense only to be willing to assert the consequent under the supposition that the antecedent is true in these cases.

Finally, two qualifications should be added to the argument above. The first caveat is that the point is not that one cannot have special cases of conditionals like (6) and (7) (cf. Bach, 2006), which at first sight don't appear to express reason relations:

> If you can lift that, I'm a monkey's uncle. (6)

> If Saddam Hussein wins the Albert Schweitzer Humanitarian Award,
> Dr. Dre will win the Nobel Prize for medicine. (7)

Rather the point is to treat it as a default reading of conditionals that they express reason relations and that such special uses have to either bracket or modify this more paradigmatic use that we are concerned with. For examples like (6) and (7) can be dealt with by noticing that apparently the speaker takes the antecedent in each case to be so preposterous that if he found himself in a position of accepting it, then he might as well accept the consequent (which expresses a proposition that he takes to be equally outrageous). However, although the line of reasoning suggested by such conditionals is obviously not one that the speaker finds worthwhile, the antecedent is strictly speaking treated as a reason for the consequent by the suggestion that accepting the antecedent should make us accept some equally absurd proposition stated in the consequent.

So far from being a counterexample to the present account, as Bach (2006) suggests, it appears that this marginal use of conditionals to express one's outrage about the absurdity of the antecedent actually exploits the fact that conditionals are normally taken to express reason relations.

To take another special case to illustrate our strategy for handling apparent counterexamples: before committing high treason by killing the king, Macbeth consults with his wife about the possible consequences. To this Lady Macbeth replies in Polanski's

filmatization "If it fails, then it fails". Of course, this is true. But on the face of it, it would seem that a redundant reason relation is articulated. However, another way of looking at the conditional is as shortcutting the discussion of what further events its potential failure would raise the probability of by in effect saying "…then come what may". According to this reading, the redundancy in the conditional achieves its effect against a background of standard uses of the conditional, whereby events that the occurrence of the antecedent would raise the probability of are under consideration.

Another apparent counterexample[41] is the so-called non-interference conditionals such as 'if it snows in July, the government will fall', where the consequent is taken to be so obvious that it will hold regardless of whether the antecedent holds. As Douven (forthcoming: 10-11) points out, one way to identify this class of conditionals is through the possibility of substituting 'if' by 'whether or not', 'regardless of whether', and sometimes by 'even if'.

One strategy in dealing with non-interference conditionals is to follow the lead of Douven and many others in accepting that a distinction between normal and special conditionals has to be accepted, because the class of conditionals is too diverse to allow for a generalization that fits all of them. Given the comprehensive, empirical classification of divergent conditionals in Declerck & Reed (2001), this may indeed be the wisest option. But even so, the view may still be retained that the relevance approach succeeds in accounting for most instances of normal conditionals.

However, I actually think that it is possible to adopt the stronger position outlined above, that a default reading of conditionals is that they express reason relations and that special uses have to either bracket or modify this more paradigmatic use (at least in relation to the standard counterexamples discussed in the philosophical literature). So returning to the non-interference conditional above, one may speculate that the 'if' in non-interference conditionals is an abbreviation of 'even if', where the latter indicates that the antecedent clause expresses something that is taken to be a reason against the consequent (either by the speaker or some other interlocutor), which the speaker holds, however, to be an *insufficient reason against* the consequent.

Accordingly, the irrelevant antecedent clause in 'if it snows in July, the government will fall' would achieve its effect by serving as a placeholder for whatever the interlocutor would like to insert with the point being that the speaker would still continue to endorse the consequent (i.e. "even if [whatever], the government will [still] fall"). In order to achieve this effect, the antecedent has to be obviously irrelevant to make it obvious that there is no particular reason why it was chosen over a different candidate. That is, what is expressed is that no matter which content is substituted for the antecedent, it would still count as an insufficient reason against the consequent.[42]

The second caveat is that the point of the argument above is not to make *every* expression of reason relations part of the semantic analysis. Under the assumption that the latter is concerned with context-invariant content that can be interpreted even on the basis of an impoverished context, it is possible that there will be cases, where propositions are connected in reason relations in ways that would normally be rejected as semantically defective, but where the epistemic situation introduces special contextual information, which introduces new reason relations. More specifically, the kind of cases I have in mind are when the agents have evidence that the true proposition is a member of a set, but they don't yet know which one it is. In such cases, eliminating possibilities will raise the probability of the remaining candidates. Hence, in this setting, propositions, that wouldn't normally count as reasons for the truth of the other propositions, will in fact raise their probability (as when a crime detective knows that the murderer was in a particular room, but doesn't yet know who it was). It is for such cases that instances of [2] have their justification. However, we cannot allow the validity of [2] to be treated as a general principle in our semantic analysis as it is only applicable, whenever the context introduces evidence that the true proposition is a member of a set without yet allowing the assignment of 'true' to any of the candidates (cf. Spohn, 2013a). Accordingly, the present approach compels us to reject the universal validity of [1] and [2] due to the fact that they don't generally preserve reason relations. At the same time, however, it still allows us to account for the special circumstances, where [2] is justified due to context-specific factors that it declares to belong to pragmatics.

3. Objective Purport and Compositionality

The purpose of this section is to sketch a strategy for dealing with the twin problems of accounting for compositionality and a sense in which there can be a factual dispute about conditionals on the basis of the relevance approach. I say that these are twin-problems, because traditionally the main strategy for dealing with compositionality has been to state the truth conditions for some proposition expressed by the linguistic content in question and then to allow that the semantic value of a more complex sentence, in which this content figures, is a function of the truth values of its elements. And, of course, once one has truth conditions of propositions, then it is a small step to hold that factual disputes concern the satisfaction of these truth conditions.

Although this has indeed been the *main* strategy, it is not the only one. Indeed, when it comes to compositionality, the main strategy of the proponents of the suppositional theory has been to argue that: (1) genuine cases of compound conditionals are rare, (2) even positions that hold that conditionals express either a truth-functional or a non-truth functional propositions [43] have their own problems when it comes to dealing with compositionality, and (3) that apparent cases of compound conditionals can be explained away by a case-to-case use of paraphrases by means of sentences that don't involve compound conditionals (cf. Edgington, 1997, 2000, 2006, Woods, 1997: ch. 6, Kölbel, 2000, and Bennett, 2003: ch. 7). To illustrate:

(i) 'if A, (if B, C)' is paraphrased as 'if (A and B), C',

(ii) ‚it is not the case that if A, then C' is paraphrased as ‚if A, non-C',

(iii) ‚if A, C and if B, D' is paraphrased as ‚if A, C. If B, D', and

(iv) '(if A, B) or (if C, D)' is taken to be virtually uninstantiated.

Yet, when it comes to dealing with the issue of factual disputes over conditionals, proponents of the suppositional theory lean towards invoking the de Finetti truth table and

saying that there can be factual disputes about the satisfaction of those truth conditions. Systematically, though, it is a bit strange that they don't also insist on using this truth table, when accounting for the compositionality of conditionals. But strategically it may be wise as Edgington (2006) points out that three-valued logics is not in a better position to avoid counterintuitive cases (regardless of how the truth tables of the other logical connectives are fixed).

A further objection that one could have towards this strategy is that it is not based on systematic principles but rather involves a free use of artistic license in selecting suitable paraphrases, when dealing with the hard cases (cf. Edgington, 1997, Kölbel, 2000).

In dealing with these twin problems on the basis of the relevance approach, one possibility is to follow Kaplan (draft) in holding that we can get a handle on expressive content by substituting the corresponding propositions that would be needed to describe it. What Kaplan is driving at is that there is an inherent limitation in truth-conditional semantics suited for descriptive content, when it comes to dealing with expressive content like 'that bastard Kaplan', 'oops', and 'ouch', which displays some state or attitude, and is more adequately explicated in terms of rules of use. However, we can make progress in applying our standard model-theoretic techniques by noticing that there is an informational equivalence between such examples and the corresponding sentences with a descriptive content. Suppose that the semantic information conveyed by descriptive content is the set of contexts at which the sentence is descriptively correct (or true), and the semantic information conveyed by expressive content is the set of contexts at which the word (or phrase) is expressively correct. It will then hold that the semantic information conveyed by 'ouch' is equal to the semantic information conveyed by 'I am in pain', and the semantic information conveyed by 'oops' is equal to the semantic information conveyed by 'I just observed a minor mishap'—or so Kaplan (draft) argues. However, he is careful to point out that they are not synonymous and they behave different logically and with respect to compositionality.

Applied to our context, we could analogously approach our twin-problems on the basis of the assumption that the expressive content of conditionals could be stated by propositions of the kind that there is a reason relation between A and C (or alternatively:

'that A is a reason for C').[44] Presumably, it is propositions of this kind that the hearer would attribute as commitments to a speaker uttering an indicative conditional.

Adopting this meta-linguistic approach [45] yields a systematic account of compositionality in that the logical connectives can now be applied to propositions of this kind in determining the content of compound conditionals. Accordingly:

(v) 'It is not the case that if A, C' gets analyzed as 'It is not the case that there is a reason relation between A and C' or 'It is not the case that A is a reason for C'.

(vi) 'If C, if A, then D' gets analyzed as 'There is a reason relation between (there being a reason relation between A and C) and D' or 'That A is a reason for C is a reason for D'.

(vii) 'If A, then C, if B' gets analyzed as 'There is a reason relation between A and there being a reason relation between B and C' or 'A is a reason for that B is a reason for C', etc..

Moreover, adopting this meta-linguistic approach allows us do justice to the perceived objective purport of assertions of conditionals, which consists in the impression that one is aiming at the truth in their assertion and that factual disagreement is possible with respect to indicative conditionals. According to this line of thought, ordinary speakers should be depicted as aiming at asserting truths about A being a reason for C, when asserting indicative conditionals, and their disagreements over such conditionals should be depicted as intended factual disputes about, whether A is *really* a reason for C.

This view connects with the work of Brandom (1994, 2010: 44-8, 104), who holds that (indicative) conditionals serve the function of making our dispositions to draw content based inferences explicit, and that one of the expressive advantages of having such a connective consists in enabling justificatory challenges that target the inferential transitions that we make implicitly. So not only can a speaker give 'if A, then C' as a justification for why he accepts 'C' in a context, where A is taken for granted. But in virtue of making his commitment to there being a reason relation between A and C explicit, his interlocutors can then subject the reason relation itself to further, critical scrutiny.

3.1 The Normative Foundation of Perceived Objective Purport

In making these points, the meta-linguistic approach is saying something about linguistic phenomenology and the semantic competence of ordinary speakers. It is then a separate issue, whether a suitable formal approach can be found, which vindicates ordinary language users in their perception of the objective purport of conditionals.

As the ranking-theoretic explication of reason relations will begin to occupy us throughout the next chapters, it will be useful at this point to consider in a purely informal way, how it would respond to the issue at hand. (Its formal introduction and the explication of reason relations in terms ranking functions representing beliefs and degrees of disbelief are postponed until section III 2.)

As far as I can see, ranking theory offers two options for reacting to the meta-linguistic approach. On the one hand, it could hold that the proposition that there is a reason relation between A and C should really be taken as shorthand for 'according to the ranking function under consideration, there is a reason relation between A and C'. In most cases, this will amount to 'I take it that there is a reason relation between A and C'. But it is also conceivable that the interlocutors may build up a mutual ranking function in the course of a conversation as common ground.[46]

Adopting this version would make the present case fully analogous to Kaplan's way of handling expressive content. In this case, conditionals would be depicted as having the expressive content of displaying aspects of the epistemic state of the speaker, which would be expressively correct just in case that they were in fact part of the corresponding ranking function. However, the downside is that this development is committed to an error-theory analogous to Mackie's error theory about ethical facts. Although it may appear to the language users *as if* they are aiming at stating objective truths about what is a reason for what in their assertions of conditionals—which is a matter that they are capable of having factual disputes about—the present approach would hold that what they were really doing was expressing a feature of their own epistemic states. Factual disputes could still be had about whether it really was the case that A is a reason for C according to a particular ranking function. But it would be factual disputes of a very different kind and disappoint

hopes of finding a rational way of adjudicating disputes between agents with conflicting ranking functions.

On the other hand, the ranking theoretic explication of reason relations could be used to take *an objectivistic approach*, which would place the content expressed by indicative conditionals on the descriptive side of Kaplan's divide between expressive and descriptive content. To explain the rationale of this approach, it is required that we briefly state the main results of Spohn's (2012: ch. 15) rather technical objectification proofs. In an attempt to show that subjective ranking functions representing degrees of disbelief are capable of possessing objective properties, Spohn notices that although a belief is an epistemic state, its truth is an objective property of it. Accordingly, if different features of ranking functions (like their expressing conditional beliefs, reason relations, beliefs about causal relations) could be characterized in terms of the beliefs that they are minimally committed to, these features would have objective truth conditions corresponding to them. More specifically, the strategy is to demonstrate that there is a one-to-one correspondence between different features of ranking functions and the minimal propositions characterizing the features in question. If the underlying ranking functions can be uniquely reconstructed from these propositions, then their truth conditions are attributed to the former. As it turns out, it can be proven that this objectification strategy will fail for the ranking-theoretic explication of the reason relation. Yet, it succeeds for the ranking-theoretic explication of causal relations as a kind of reason relation conditioned on the actual history.

In reflecting on the significance of this result, Spohn (2012: 369) says that:

> Now in our subjectivistic approach (direct) causes simply were a particular kind of conditional reasons, and Sections 15.4–15.5 proved that if we assume a specific temporal and logical form for these conditional reasons, we can place them in a one-one correspondence with objective material implications. So, it seems the causal relation is just the well-formed objectifiable part of our much richer and more disorderly reason relation. In other words, if we want to objectify our inductive strategies, if we want to align our dynamics of belief to the real world, we have to attend to causation, to the objectifiable part of our reasons. This is what the notion of causation is for.

Accordingly, the success in showing that the ranking-theoretic explication of causal relations can be brought into a one-to-one correspondence with material implications, introduces the prospect of striving for objectivity in our reason relations by aligning them with the objectifiable causal relations.

Applied to the meta-linguistic approach, this result would vindicate the perceived objective purport of assertions of indicative conditionals to the extent that truth conditions of these causal relations could be ascribed to the propositions stating reason relations. Accordingly, ordinary language users would be justified in their perception that they were aiming at the truth in asserting (contingent)[47] indicative conditionals, and that it is possible to have factual disputes about them, to the extent that: (a) they were thereby aiming at stating reason relations that were capable of being identified with causal relations, and (b) such discussions are depicted as being discussion about, whether the reason relation expressed by a given indicative can be considered a causal relation. So if, for instance, there is a dispute about the indicative 'If the glass is dropped, then it will break', this dispute can be reconstructed as a factual dispute about, whether the corresponding counterfactuals expressing causal relations would be true (e.g. 'If the glass had been dropped, it would have broken') as the objectification strategy allows us to assign truth conditions to counterfactuals (cf. Spohn, forthcoming).[48]

However, no matter whether the former subjectivistic or the latter objectivistic approach is preferred, it holds that they are attempts of *justifying* the perceived objective purport in asserting indicative conditionals using the resources of ranking theory by *regimenting* factual disputes over indicative conditionals as either concerning features of the agent's own ranking functions or objective, causal relations. In neither case should we view these regimentations as descriptions of the linguistic competence of ordinary speakers. For the purposes of the latter, we need not go beyond the meta-linguistic approach that we started out with. That this is so can be seen by the fact that the subjectivistic approach ended up being committed to an error-theory about the objective purport of assertions of conditionals, and that the objectivistic approach ended up relying on some very technical proofs that employed the identification of minimal propositions as part of an elaborate proof strategy.

Hence, as compositionality deals with an aspect of linguistic competence neither the subjectivistic nor the objectivistic truth-conditions should be used to account for the semantic content that ordinary language users associate with compound conditionals.

3.2 Comparative Remarks

When one inspects the various truth tables that have been put forward for indicative conditionals, it turns out that they have difficulties with accommodating the natural idea of assertions of conditionals as aiming at the truth. This is a bit of an embarrassment as this notion is standardly invoked in analyzing unconditional assertions and no reasons have been provided (that I am aware of) of why it shouldn't be of equal use in the case of assertions of conditionals.

To illustrate, it is not generally the case that we aim at stating the truth that A & C is the case, when asserting an indicative conditional as the point of conditional assertions is that one can remain uncommitted about the truth of C by being uncommitted by the truth of A. Hence, the de Finetti table is not adequate to capture this idea of aiming at the truth. Similarly, it is not generally the case that we aim at stating the truth that ~A v C, when asserting an indicative conditional as we could just have made an unconditional assertion about non-A, if this was our belief. Hence, the material implication is not fit to capturing the idea of aiming at the truth in the assertion of an indicative conditional either. In contrast, the meta-linguistic approach was able to capture this idea by holding that in asserting an indicative conditional, the speaker is aiming at the truth of the proposition that there is a reason relation between A and C. (Of course, the subsequent analysis then revealed that this attitude is only justified on the basis of ranking theory, if either the regimentation suggested by the subjectivistic or the objectivistic line is implemented. But this is a different matter. Now a point is made about the semantic analysis being able to account for the *appearance* of aiming at the truth, when asserting a conditional and it is pointed out that accounts based on the de Finetti table or the material implication encounter difficulties even in accomplishing that.)

The next comparative point concerns Edgington's remarks that: "[c]ompounds of conditionals are hard: much harder than one would expect if conditionals have truth

conditions" (2000: 115) and that "no theory has an intuitively adequate account of compounds of conditionals" (2006). In the same context, she tends to make remarks about the relative rareness of different compound constructions and of how difficult it is to process compounds like "If Kripke was there if Strawson was there, then Anscombe was there". In response, I want to point out that an impression of the relative rareness of different compound constructions—and a small handful of hard cases—should not be allowed to act as a substitute for a corpus analysis for their relative frequency and the latter's representativeness among the compound found. Only the latter should be taken as decisive in settling this matter. Yet, this debate has been allowed to go on for over 20 years without making use of this standard tool in linguistics.

The second point that I want to make is that the approach to compositionality suggested by the meta-linguistic approach should make us suspect that there will be a higher frequency of compound conditionals in conversations and texts, where the interlocutors get sophisticated about composing a dialectical structure. To illustrate, one can find the three following gems naturally occurring in an unpublished argumentative text by Kaplan (which is not on the subject of conditionals):

(i) IF Grice is right about the descriptive content of the premise in Argument 5, then UNLESS it is not valid, then I am wrong and Grice is right about logic, OR information conveyed expressively cannot be converted into information conveyed descriptively. (draft: 26) [If A, then, unless non-B, C and D, or non-E.]

(ii) Analogously, IF the descriptive content of the second premise of Argument 6 includes the expressive content in the conclusion, then again, IF the argument is not valid, then I am wrong about logic, UNLESS information conveyed descriptively cannot be converted into information conveyed expressively. (ibid) [If A, then if non-B, then C, unless non-E.]

(iii) If I am correct about parts of language being marked to *display* respect (or disrespect), then the use of such language, even if thought to be insincere, is *respectful behavior*, and should produce an affective response in its own right. (draft: 31) [If A, then C, even if B, and D.]

Until now the proponents of the suppositional theory have not come up with any paraphrase of examples like this that I am aware of. But they can easily be accounted for on the basis of present approach. By making the following two assumptions, the rendering below is possible: (1) 'unless' serves the specific role of stating a condition that would undermine the reason relation (which we will begin to call a 'disabler' in the next chapter), and (2) 'C even if A' serves the specific role of indicating that A is taken to be a reason against C in the context of the conversation (either by the speaker or by his interlocutors), yet the speaker holds that A is an insufficient reason against C:

(i) Either there is a reason relation between A and (C and D), which would be undermined by non-B, or E is not the case.

(ii) There is a reason relation between A and there being a reason relation between non-B and C, which would be undermined by non-E.

(iii) There is a reason relation between A and (C and D), whereby B is an insufficient reason against C.

Returning to "If Kripke was there if Strawson was there, then Anscombe was there", it can be rendered as: there is a reason relation between (there being a reason relation between Kripke being there and Strawson being there) and Anscombe being there. What is a bit odd about this example is why the speaker should assume there to be such a relationship. But suppose that Anscombe believed that Kripke was attending the conference, and that it is really Strawson that she wanted to see. If then Anscombe assumed that Kripke's presence raised the probability of Strawson's presence, then this might in turn have raised the probability of Anscombe's presence.

Or to take another example with the same syntactical form: 'If the glass broke, if it was dropped, then it was fragile'. In this case we have no problem with supposing that the probability of the glass being fragile is raised in case the probability of the glass being broken is raised by its being dropped.

Furthermore, Edgington (1995, 2006) has been claiming for some time by now that 'Either (if A, C) or (if B, D)' is virtually uninstantiated, whenever the speaker is open about either disjunct. However, it is not hard to cook up examples, where a speaker could take it that there is either a reason relation between A and C or a reason relation between B and D and be open about either disjunct. Suppose that a student is preparing for an exam and says to one of her fellow students that "Either, if the question about Kant comes up, the correct answer is rationalism, or, if the question about Hume comes up, the correct answer is empiricism. I can't remember which" as a way of reporting what the teacher said. Although this way of talking is perhaps a bit convoluted, it is not hard to understand what is being communicated.

So far our account of compositionality seems to be in a good shape as it has a systematic approach that applies even to the hard cases. Another systematic approach to compositionality is the truth-functional account. However, it faces problems in dealing with 'If the glass broke, if it was dropped, then it was fragile' as Edgington (2006) points out. The reason is that the conditional embedded in the antecedent is true, if the glass was not dropped according to this analysis. Yet, if it simultaneously holds that the glass is not fragile, then we will have a case with a true antecedent and a false consequent, which renders the compound conditional false. However, it seems that the conditional is capable of being true (and acceptable) under those circumstances.[49]

A second counterexample to the truth-functional account that Edgington (2006) points to is the case of negations of conditionals. According to the truth table for the material implication, its negation should be equivalent to 'A & non-C'. As a result, it faces problems in dealing adequately with cases like saying of an unseen geometric figure that 'It's not the case that if it's a pentagon, it has six sides'. The reason is that this compound conditional will then be false, if it turns out not to be a pentagon. Yet, it seems wrong that the issue of whether the unseen geometric figure is in fact a pentagon should decide the issue of whether it's being a pentagon is a reason for taking it to have six sides.[50]

In contrast, on our analysis, the compound denies that there is a reason relation between the geometric figure being a pentagon and it's having six sides, which seems exactly right. In general, one can use the negation of a conditional to deny that A is reason

for C either because: (1) one withholds judgment, (2) one thinks that they are irrelevant to one another, or (3) one takes A to be a reason against C. In this case, it is surely (3) that is the intended reading.

On the basis of this comparative discussion, I conclude that the present approach to compositionality is in a good shape. However, one technical challenge is that it made use of the idea of reason relations as themselves entering in reason relations. Yet, the ranking theoretic explication of the reason relation that we will encounter in the next chapter is only defined over propositions. As a result, the reason relation cannot itself enter as a relatum in other reason relations on this explication. But perhaps one solution to this problem would be to treat 'the proposition that there is a reason relation between A and C' as the requisite relatum.

It might appear as if this in turn would require use of second order ranking functions representing the agent's beliefs about his own reason relations, but that such an impression would be mistaken is easily seen. In order to assess whether A is a reason for C, the agent merely has to assess whether $P(C|A) > P(C|\overline{A})$. However, in order to make such an assessment, the agent need not access his own beliefs about A as all he is being required to do is to estimate whether the probability of C would be greater under the supposition that A is the case than under the supposition that A is not the case. It is thus possible for him to hold that 'Russia has just launched a missile attack against Washington DC' is a reason for 'The Third World War has begun' without actually believing that Russia has done such a reckless thing.

Similarly, once we consider the case of 'a proposition about a reason relation between A and B' being a reason for C, the agent only needs to assess whether the probability of C would be greater under the supposition that the reason relation between A and B obtains than under the supposition that it doesn't obtain. At no point does the agent need to access his own ranking functions to find out whether he actually takes A to be a reason for B. Driving home on a rainy day to work, the agent may thus entertain the thought that 'if the car starts slipping, if I make a sudden turn, then the road is too wet to continue driving' without actually believing that the car will start slipping if he makes a sudden turn as he takes the amount of rain fallen to be negligible. Hence, the present

account of compositionality steers free of technical difficulties associated with second order ranking functions in spite of the fact that it makes use of the idea that propositions about reason relations can themselves enter as relata in further reason relations. (However, there may be further technical difficulties associated with introducing propositions about reason relations as members of the algebra that our informal treatment has not yet produced a solution to.)

---

[29] Acknowledgement: this chapter profited by discussion and comments made by Wolfgang Spohn, Arno Goebel, and the participants at Thomas Müller's colloquium at the University of Konstanz.

[30] Reference: in Pfeifer & Kleiter (2011) corroborating evidence was found that it is not just the tainted intuitions of theoreticians that lead to the rejection of these argument schemes but that ordinary subjects reject them as well.

[31] Reference: Gauker (2005: ch. 3), Douven (forthcoming: ch. 2).

[32] Reference: Lycan (2001: ch. 4) makes much of this point in his criticism of the suppositional theory.

[33] Reference and explication of de Finetti table: some attempts of specifying a sense in which conditionals have trivial truth values include Blackburn (1986), Bennett (2003: ch. 8), Edgington (2003), and Politzer, Over, & Baratgin (2010). It seems to me that one difficulty with adopting this approach, and holding that the truth values of conditionals is given by the de Finetti table, is that the probability of the proposition thus defined is equal to $P(A \cap C)$ and not to the desired $P(C|A)$. But perhaps it could be argued that once we begin treating the false antecedent cases as void, we need to renormalize, so that we still get a probability distribution, where the probabilities of complementary events sum up to one. If that is the idea, we would indeed get that $P(\text{if } A, C) = P(A \cap C)/P(A) = P(C|A)$ as what we have done is in effect to conditionalize on A.

[34] Extension: this problem is actually quite severe for the suppositional theory of conditionals, if Kölbel (2000) is right that denying that indicatives have truth conditions commits the proponents of the suppositional theory to holding that conditionals are to be understood as a syntactical device that functions as a complex force indicator indicating that the consequent is only asserted conditionally on the antecedent being asserted.

[35] Caveat: in light of the comparative advantages of p-validity outlined in section 1.2, it should be noted that the logic that be developed on the basis of this notion of probability difference making remains to a large extent unexplored territory. It would seem important for the comparative discussion that more

progress is made with respect to this issue. In Spohn (2012, ch. 6, forthcoming) and in Douven (forthcoming: ch. 5) some important first steps are taken, however.

[36] Answering a possible objection: if the reader is unfamiliar with guinea pigs (in German: Meerschweinchen), it may appear as if their lack of tails is a fact that goes beyond semantic competence. To avoid such worries, the reader should feel free to substitute the following example instead: 'If you tickle a cat under its wing, then its eyes will fall out'.

[37] Caveat concerning analyticity: as it happens, the notion of synonymy, or of analytical truths that hold good in virtue of the meaning of words, has itself become disputed territory in philosophy after the publication of Quine's famous paper 'Two Dogmas of Empiricism'. However, at this point I am following Williamson (2008: 50-51) in holding that in spite of problems of finding a non-circular definition of these notions, sacking them seems like an overly strong response in the context of a commitment to naturalism, insofar as linguists are happy to work with these notions.

[38] References: the list up until this point is called *speech act adverbials* in Bach (1997) and *utterance modifiers* in Bach (2006). The items after it have been added by me. Whether the grammatical label is still adequate I am unsure. But it seems that they perform a similar role in making the dialectical structure of the text or conversation explicit. In Blackmore (2004), expressions like those included on my list are called *discourse markers*. But as she also points out, this class is a mixed bag in linguistics, whose extension differs from author to author. Some of the examples she considers have nothing to do with argument structure and they would thus not be covered by the present proposal. Moreover, in chapter 4 she highlights complications to the analysis such as that 'but', 'nevertheless', and 'however' are not always interchangeable although they all appear to express the same reason against relation.

[39] Reference: in Blackmore (2004: ch. 4) one finds a defense of the semantic analysis that 'but' plays the role of denying implicit or explicit assumptions against apparent counterexamples. This is not quite the same as saying that there is a contrastive relationship between the conjuncts, because the assumption denied may remain implicit, but it goes in the same direction.

[40] Reference: an important antecedent in the literature with respect to this topic is Merin (1999), who offers a decision theoretic semantics of 'or', 'not', 'but', 'even', and 'also', where the formal explication of relevance in terms of probability difference making plays a crucial role.

[41] Caveat concerning other conditional speech acts than conditional assertions: here I am not including so-called *biscuits conditionals* (e.g. 'there are biscuits on the table in the kitchen, if you want any') as a possible exception as they seem to take the form of a conditional tip and we have been concerned with assertions of conditionals. Similarly, I am not including sentences like 'If you ask me, he's Italian' (Dancygier, 2003: 312) as a possible counterexample for the same reason. In this case, the speaker is indicating his disposition to react to a request for information and would presumably be prepared to

assert the consequent regardless of whether the antecedent holds (that is, regardless of whether the speech act is performed of requesting him for information). Of course, in the end it would be desirable to have a general theory that could cover all conditional speech acts, but for present purposes we are pursuing a more modest goal. If one were to pursue the more ambitious goal, then it appears that the relevance approach would have to be supplied with an account of conditionals as specifying conditions under which some deontic state of affairs hold (e.g. a promise being issued) to account for deontic conditionals. A classification of this type of conditionals in linguistics is 'purely case-specifying conditionals', where the antecedent clause specifies the condition under which the speech act in the consequent clause is felicitously addressed to the hearer (ibid: 319).

[42] Further potential counterexamples: a different non-interference conditional discussed in Douven (forthcoming: ch. 4) is 'If Reagan is bald, no one in the press knows it'. This case differs from the one dealt with in the text in that here the antecedent is not irrelevant for the consequent. However, the hypothesis that 'if' is an abbreviated version of 'even if' still seems applicable.

In relation to the example 'If Reagan worked for the KGB, we will never find any evidence for that', Douven (forthcoming: ch. 4) points out that the 'whether or not' paraphrase may not be applicable in contexts, where it is preceded by 'If Regan worked for the CIA, then, sooner or later, we will come to have evidence for that. On the other hand,…'. However, it seems that the consequent can be paraphrased in terms of 'we *still* won't find any evidence for that', which seems to be the mark of a concessive conditional. So it seems that this conditional would also be a candidate for the *insufficient reason against* interpretation. Yet, Douven suggest that the 'even if' paraphrase would fail under these circumstances. If this is so, then perhaps it is due to the embedding under 'on the other hand'. It seems that the function of 'on the other hand' in this example is to indicate that there is a contrast between the antecedent being *a (sufficient) reason for* the consequent in the first conditional and the relationship between the antecedent and the consequent in the second conditional (where the antecedent is *an insufficient reason against* the consequent).

[43] Explication of truth-functional and caveat: the distinction between the two is that between truth-tables, where the truth value of the conditional is a function of the truth values of the components (e.g. the material implication) and truth-tables, where the truth values of the components leave open the truth value of the conditional (e.g. possible world semantics holds that when the antecedent is false, the conditional may either be true or false depending on whether the consequent is true in the nearest possible world in which the antecedent is true) (Edgington, 2006).

[44] Commentary to the analogy: as always, the analogy is far from perfect and there are disanalogous aspects as well. In particular, expressives like 'that bastard' behave differently with respect to compositionality than our propositions about reason relations. As Kaplan (draft) points out, although

embedding in the antecedent clause in a conditional is normally taken to bracket the assertive force of the proposition, the derogatory effect of 'that bastard Kaplan' is still achieved in 'If that bastard Kaplan gets promoted, then…'. However, the analogy will still be useful to the extent that it allows us to transpose a solution from one domain to solve a problem in a different domain.

[45] Superficial similarity with Goodman's account: the name is chosen due to a superficial similarity with the approach to counterfactuals advanced in Goodman (1991, [1947]), which adopts the following analysis: "There is some true proposition Support such that: ~ (if A, ~ Support) and (A & Support & laws) entails C" (Bennett, 2003: 308, notation modified, NSO). In contrast, the present view only holds that A is a sufficient reason for C, which raises it above the threshold of belief (cf. chapter III), and whereas counterfactuals are usually taken to express causal relations, we initially leave it an open question whether the speaker takes the reason relations expressed by indicative conditionals to be based on causal relations as well.

[46] Acknowledgement: this idea emerged in a conversation with Arno Goebel.

[47] On non-contingent indicative conditionals: notice that there is no problem in providing an objective basis for our non-contingent indicative conditionals (e.g. in mathematics) as the notion of deductive reasons that can be explicated on the basis of the subset relation is not relativized to doxastic states (cf. Spohn, 2012: 109).

[48] Extension concerning regularities: however, insofar as it argued by Spohn that true regularities in nature is the objective counterpart of his epistemic specification of causal relations, it could be argued that one would be able to align one's reason relations to the true regularities in nature even in the case of common cause scenarios. If so, then the antecedent need not be a cause for the consequent and the objectivistic approach would have to be generalized to dealing with true regularities in nature as opposed to be dealing merely with causal relations.

[49] Potential objection: to be sure, Douven (forthcoming: 40) points out that the sentence could in principle be construed as 'If {the cup was dropped}, then [if it broke, it was fragile]' instead of as 'If {the cup broke if it was dropped}, then [it was fragile]'. But the latter appears to be the most natural reading.

[50] Potential objection: here Douven (forthcoming: 49-50) points out that the proponents of the truth-functional account could follow the suppositional theory in holding that negations of conditionals are to be read as negations of their consequents by arguing that surface grammar is misleading. They would thereby avoid this second problem. However, once they begin making use of the same artistic license as the proponents of the suppositional theory in dealing with apparent counterexamples, they lose the principal virtue of their account: its systematicity (i.e. that logical connectives are applied directly to the truth conditions attributed to indicatives).

III

Making Ranking Theory Useful for Experimental Psychology[51]

Abstract: The idea from chapter I about expanding the hypothesis space of serious possibilities was illustrated in chapter II by the introduction of the relevance approach as a contender in psychology of reasoning. To the extent that the argument given was successful, this theoretical possibility should have a high prior as compared to the other candidates currently under consideration in psychology.[52] Normally, philosophers would have left it at that (perhaps after making it formally precise and establishing its compatibility with the existing data). But the point of the argument in chapter I was precisely that we cannot rest content with having introduced more uncertainty by expanding the hypothesis space, without giving any directions of how the experimentalists are to reach an empirically grounded decision among the now increased number of candidates. For this reason, the next two chapters are devoted to implementing the methodological recommendations from chapter I with respect to the relevance approach to conditionals as it is advanced in Spohn (2013a, forthcoming). A first, important step is taken in this chapter, when a statistical model called logistic regression is used to extend ranking theory in a way, which makes it possible to derive quantitative predictions for psychological experiments. Finally, an appendix has been added, which puts forward an alternative taxonomy of reason relations.

1. Introduction

Ranking theory has been developed into a comprehensive, formal epistemology in over 600 pages in Spohn (2012) that is able to provide a normative account of the dynamics of beliefs and non-monotonic reasoning. In fact, its originator claims that the study of ranking functions is *the* study of beliefs (Spohn, 2009), that ranking theory delivers *the* dynamic laws of beliefs, and that it is *the* legitimate sister of probability theory (Spohn, 2012: xii). Recently, the theory has been extended to lay the epistemological basis for a semantics of conditionals (Spohn, 2013a) that makes the idea of a relevance approach to conditionals advanced in chapter II more precise.

Ranking theory has already been received in the AI community (cf. Goldszmidt & Pearl, 1996), but its application in psychology has still to come. I already know of several psychologists, who have shown an interest in testing it. But they have been unable to derive clear, experimentally distinguishable predictions from it. As we shall see, the theory of conditionals presented in Spohn (2013a) provides some qualitative predictions. But it is not clear how to turn these into precise, quantitative predictions. The extension of ranking theory to be presented in section 3 improves the situation. Yet, it violates a formally powerful translation of probabilities into ranking functions. Section 2 will therefore challenge this translation.

2. Arguments against the Infinitesimal Translation

2.1 Introducing Ranking Theory

Before we dwell into these topics, it will serve our purposes, if we first have a simple presentation of ranking theory, which can later be extended whenever needed.

Ranking theory is built up on a metrics of beliefs, which quantifies a grading of *disbelief* expressed by *negative* ranking functions, κ. The object of our degrees of disbelief is taken to be propositions (i.e. the content shared by sentences of different languages). To formally represent propositions, ranking theory follows possible world semantics in

representing propositions as sets of "possible worlds" or possibilities. So to state ranking theory, first a non-empty set, W, of mutually exclusive and jointly exhaustive possibilities is assumed. Next an algebra, $\mathcal{A}$, of subsets of W is formed that is closed under logical operations. This collection of subsets of W represents all possible propositions. Doxastic attitudes such as *believing* and *disbelieving* propositions can then be represented by functions that are defined over $\mathcal{A}$. Accordingly, negative ranking functions expressing an agent's degrees of disbelief can now be defined as follows:

> *Definition 1*: Let $\mathcal{A}$ be an algebra over W. Then κ is a *negative ranking function* for $\mathcal{A}$ iff κ is a function from $\mathcal{A}$ into $N \cup \{\infty\}$, the set of natural numbers plus infinity, such that for all $A, B \in \mathcal{A}$:

$$\kappa(W) = 0 \text{ and } \kappa(\emptyset) = \infty \tag{1}$$

$$\kappa(A \cup B) = \min\{\kappa(A), \kappa(B)\} \tag{2}$$

where $\kappa(A)$ is called the *negative rank* of A. From the above it follows that:

$$\kappa(A) = 0 \text{ or } \kappa(\overline{A}) = 0 \text{ or both} \tag{3}$$

If $\kappa(A) < \infty$, then the *conditional rank* of B given A is defined as follows:

$$\kappa(B \mid A) = \kappa(A \cap B) - \kappa(A) \tag{4}$$

Since negative ranks are said to represent degrees of disbelief, $\kappa(A) = 0$ represents that A is *not disbelieved*. When $\kappa(A)$ assigns a value of $n > 0$ to A, then A is said to be disbelieved to the n^{th} degree. *Doxastic indifference* is represented by neither disbelieving A nor ~A, i.e. $\kappa(A) = \kappa(\overline{A}) = 0$, and *belief in A* is represented in an indirect way by disbelief in ~A, i.e. $\kappa(\overline{A}) > 0$.[53]

On this basis, *positive ranking functions* representing degrees of beliefs can be defined for A by:

$$\beta(A) = \kappa(\overline{A}) \tag{5}$$

Positive ranking functions can then be axiomatized by translating (1), (2) and (4) into their positive equivalents:

$$\beta(W) = \infty \text{ and } \beta(\emptyset) = 0 \tag{6}$$

$$\beta(A \cap B) = \min\{\beta(A), \beta(B)\} \tag{7}$$

$$\beta(B \mid A) = \beta(\overline{A} \cup B) - \beta(\overline{A}) \tag{8}$$

Moreover, it is possible to define two-sided ranking functions for A that combine the gradings of disbelief and belief into one function:

$$\tau(A) = \beta(A) - \kappa(A) = \kappa(\overline{A}) - \kappa(A) \tag{9}$$

$$\tau(B \mid A) = \beta(B \mid A) - \kappa(B \mid A) = \kappa(\overline{B} \mid A) - \kappa(B \mid A) \tag{10}$$

When one considers the further theorems that can be proved on the basis of these axioms, it becomes apparent that there is a deep parallel between negative ranking functions and probability distribution functions as exhibited in table 1 below (Spohn, 2012: ch. 5):

Table 1, Comparison between the Probability Calculus and Ranking Theory

Probability Calculus	*Ranking Theory*
$P(A \cap B) = P(A) \cdot P(B \mid A)$	$\kappa(A \cap B) = \kappa(A) + \kappa(B \mid A)$
$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$	$\kappa(A \mid B) = \kappa(B \mid A) + \kappa(A) - \kappa(B)$
$P(B) = \sum_{i}^{n} P(B \mid A_i) \cdot P(A_i)$	$\kappa(B) = \min_{i \leq n}[\kappa(B \mid A_i) + \kappa(A_i)]$

This is no accident. As Spohn (2009) points out, probabilities can be translated into negative ranks. By applying the translation manual below, one is *almost* sure to obtain a ranking theorem from any probabilistic theorem:

> There is obviously a simple translation of probability into ranking theory: translate the sum of probabilities into the minimum of ranks, the product of probabilities into the sum of ranks, and the quotient of probabilities into the difference of ranks. Thereby, the probabilistic law of additivity turns into the law of disjunction, the probabilistic law of multiplication into the law of conjunction (for negative ranks), and the definition of conditional probabilities into the definition of conditional ranks. If the basic axioms and definitions are thus translated, then it is small wonder that the translation generalizes; take any probabilistic theorem, apply the above translation to it, and you almost surely get a ranking theorem. (p. 209)

If negative ranking functions are treated as the logarithms of probabilities with a base, $a \in (0,1)$, the translation of products and quotients of probabilities as the sum and difference of ranking functions is captured. However, Spohn (ibid., 2012: 203) points out that if the sum of probabilities is to be translated into the minimum of ranks, the logarithmic base has to be infinitesimal. So for purposes of theoretical unification the latter translation seems superior. Yet, for psychological purposes this translation is deeply problematic as I will now go on to argue in the next sections.

Of course, it would also have been possible to apply ranking theory directly to psychological experiments,[54] instead of taking the indirect route suggested below of first translating negative ranking functions into probabilities and *then* applying the theory to the data. However, the reason why the latter, more conservative strategy is presented here is that it has the advantage of enabling the kind of direct comparison with existing findings and experimental paradigms presented in chapter IV.

2.2 Implications for the Probability Scale

The first problem for the infinitesimal translation is that it would require that subjects had all their degrees of disbelief in A expressed in probabilities from 0 to $0 + \varepsilon$, where ε is an infinitesimal quantity that is bigger than but arbitrarily close to zero. To illustrate, we can work through one of Spohn's examples using a logarithmic base of $1 \cdot 10^{-6}$ as an approximation (remembering that all the examples become more extreme the closer we

move the logarithmic base to 0). Illustrating ranking theory using the famous Tweety case, which exemplifies non-monotonic reasoning by showing that our degree of belief in that a bird (B), Tweety, flies (F) changes after we discover that it is a penguin (P), Spohn (2009) describes the doxastic state by assigning the values shown in table 2:

Table 2, Tweety Example

κ	$B \& \overline{P}$	$B \& P$	$\overline{B} \& \overline{P}$	$\overline{B} \& P$
F	0	5	0	25
$\overline{F}$	2	1	0	21

Using our approximation, this would yield the probabilities shown in table 3:

Table 3, Translation into Probabilities using the Approximation

p	$B \& \overline{P}$	$B \& P$	$\overline{B} \& \overline{P}$	$\overline{B} \& P$
F	$?^{55}$	$1. \cdot 10^{-30}$	?	$1. \cdot 10^{-150}$
$\overline{F}$	$1. \cdot 10^{-12}$	$1. \cdot 10^{-6}$	?	$1. \cdot 10^{-126}$

If we continue with our approximation, we have an interval of (0.001, 1], where values are obtained that would yield a negative rank of zero if rounded off. In order to have a case of doxastic indifference, both ~A and A would need to receive a rank of zero. Since $P(A) + P(\overline{A}) = 1$, we need $P(A) \leq 0.001$ to express disbelief in A (and belief in ~A), $P(A) \geq 0.999$ to express disbelief in ~A (and belief in A), and $0.001 < P(A) < 0.999$ ends up expressing doxastic indifference. Qualitatively, it is very difficult to make sense of these values.

Now it may be that there are examples like the one that Goldszmidt & Pearl (1996) describe in the following quote, where we have to reason about rare events, which they use to motivate the infinitesimal translation with. However, they can hardly be taken to be representative for our beliefs in general:

> The uncertainty encountered in common sense reasoning fluctuates over an extremely wide range. For example, the probability that the new book on my desk is about astrology is less than one in a million. However, if I open the wrappings and see a Zodiac sign, the

> probability comes close to 1, say 0.999. Intelligent agents are expected to reason with such eventualities and to produce explanations and actions whenever these occur. (57-58)

Moreover, even a probability that is less than one in a million is a crude approximation to a probability that is infinitely close to zero.

So when Goldszmidt & Pearl (1996) go on to suggest that table 4 may be an appropriate verbal translation of our degrees of beliefs, we have every right, I think, to remain skeptical:

Table 4, Goldszmidt & Pearl's (1996) verbal-numerical scale

$P(A) = \varepsilon^0$	A and $\overline{A}$ are believable	$\kappa(A) = 0$
$P(A) = \varepsilon^1$	$\overline{A}$ is believed	$\kappa(A) = 1$
$P(A) = \varepsilon^2$	$\overline{A}$ is strongly believed	$\kappa(A) = 2$
$P(A) = \varepsilon^3$	$\overline{A}$ is very strongly believed	$\kappa(A) = 3$
...	…	…

Pfeifer (2002) shares my skepticism about the psychological utility of investigating degrees of beliefs that are expressed at this end of the probability scale for the vast majority of our beliefs when he says:

> *Infinitesimal* probability semantics requires a conditional probability $P(\beta|\alpha)$ infinitesimally close to 1. In daily life such extremely high probabilities can be observed very seldom. Exceptions are tossing a coin such that it lands neither head nor tail side up but upright on its edge, or if someone takes a flight, she assumes that the probability to get involved in a plane crash is infinitesimal small. (17)

For his own experiments he has thus made the more convenient choice of a non-infinitesimal probability semantics over the prospect of having to explain to his participants that they have to express their degrees of beliefs in probabilities that are infinitely close to 1. In the end, it is of course an empirical question which fineness of grain verbal descriptions of the participants' beliefs are given in. However, in no psychological experiment that I know of has it been found useful to test, whether the participants are able to discriminate degrees of disbeliefs infinitely close to zero, and it is a standard

procedure to use scales that spread out more evenly across the probability scale. This suggests that the infinitesimal translation is detrimental to any use of ranking theory in psychology.

2.3 Ramifications for the Applications of Ranking Theory

As expected, the general problem outlined in section 2.2 has ramifications for the various applications of ranking theory exhibited in Spohn (2012). In this subsection we will briefly look at one example, which at the same time serves expository purposes. But more could easily be found.

In our example we turn to Spohn's (2012: ch. 6) epistemic notion of relevance, which is given a pivotal role in his semantics of conditionals (2013a) and account of causation (2012: ch. 14). Inspired by the notion of statistical dependency and independency, Spohn defines relevance as follows:

A is positively relevant to C iff	$\tau(C \mid A) > \tau(C \mid \overline{A})$	(11)
A is irrelevant to C iff	$\tau(C \mid A) = \tau(C \mid \overline{A})$	(12)
A is negatively relevant to C iff	$\tau(C \mid A) < \tau(C \mid \overline{A})$	(13)

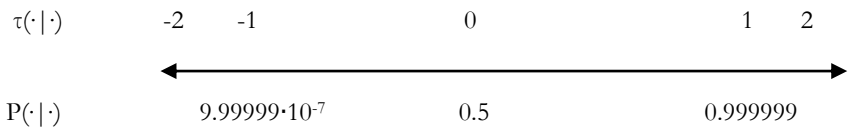
Furthermore, Spohn holds that this notion can be used to analyze the notion of reasons by holding that A is a reason *for* C iff (11) holds and a reason *against* C iff (13) holds. He is then able to use this notion of reasons to analyze four types of reason relations:

Supererogatory reason	$\tau(C \mid A) > \tau(C \mid \overline{A}) > 0$	(11a)
Sufficient reason	$\tau(C \mid A) > 0 \geq \tau(C \mid \overline{A})$	(11b)
Necessary reason	$\tau(C \mid A) \geq 0 > \tau(C \mid \overline{A})$	(11c)
Insufficient reason	$0 > \tau(C \mid A) > \tau(C \mid \overline{A})$	(11d)

In Spohn (2013a) these constructions are put to use when it is suggested that conditionals have a range of expressive functions that go beyond the Ramsey test such as expressing the reason relations described by (11)-(13) above.[56] Whereas these inequalities

focus our attention on the extent to which the antecedent is rank (or probability) raising for the consequent, the Ramsey test merely consists in adding the antecedent to our knowledge base and evaluating the probability of the consequent on this basis. Hence, these constructions hold the promise of making the relevance approach precise that we encountered in chapter II.

This all sounds terribly plausible—until the infinitesimal translation is employed. To illustrate consider first the scale of two-sided ranking functions generated by using our approximation:



If we now suppose that $\tau(C \mid \overline{A}) = 1$, and continue to use our approximation, then $P(C \mid A) > 0.999999$ is required for A to be positively relevant for C. Moreover, $P(C \mid A) < 9.99999 \cdot 10^{-7}$ is required for A to be negatively relevant for C whenever $\tau(C \mid \overline{A}) = -1$. That is, whenever $\tau(C \mid \overline{A}) = 1$, A can only be a reason *for* C whenever $P(C \mid A) > 0.999999$, and whenever $\tau(C \mid \overline{A}) = -1$ A can only be a reason *against* C whenever $P(C \mid A) < 9.99999 \cdot 10^{-7}$. Yet, $\tau(C \mid \overline{A})$ was set to the lowest possible rank above zero in the first case and to the highest possible rank below zero in the second!

Clearly it is unrealistic to assume that inductive reasons come with this degree of certainty. As a result, for inductive reasons the notion of positive relevance is only interesting whenever $\tau(C \mid \overline{A}) \leq 0$, and the notion of negative relevance is only interesting whenever $\tau(C \mid \overline{A}) \geq 0$. Hence, the notion of A being an inductive, supererogatory reason for or against C becomes problematic despite its conceptual elegance. Moreover, if $\tau(C \mid A)$ is set to 1, then A can only be a reason *against* C if $P(C \mid \overline{A}) > 0.999999$, and if $\tau(C \mid A)$ is set to -1, then A can only be a reason *for* C if $P(C \mid \overline{A}) < 9.99999 \cdot 10^{-7}$. So the notion of A being an inductive, insufficient reason for or against C becomes problematic.

Finally, we get odd results for when A is either a sufficient or a necessary reason for or against C across the interval [-2, 2]. The underlying problem is that the distances between -1 & 0 and 0 & 1 are incredibly large, whereas the distances between -1 & -2 and 1 & 2 are almost disappearing. Again it is difficult qualitatively to make sense of the notion that when one climbs up the ladder of sufficient reasons for C, then at the first step A has to raise the probability of C with almost 0.5 and at every further step A merely has to raise the probability of C with a number that is extremely close to zero. Correspondingly, we get similar odd results for the other types of reason relations mentioned that fall within this interval.

The diagnosis is that this grading is simply too coarse in the interval of probabilities between $9.99999 \cdot 10^{-7}$ and 0.999999 to be useful. Yet, it is arguably in this interval that all the action takes place in our everyday inductive reasoning (and remember: the better the approximation, the worse it gets). Hence, these observations severely challenge the usefulness of what is otherwise an elegant account of inductive reasons. However, the account of reasons in chapter six is so central to the project in *The Laws of Beliefs* that it would be a disaster, if it were to take a hit.

2.4 Dilemma

Taken together, the objections from sections 2.2 and 2.3 against the infinitesimal translation pose a dilemma for ranking theory: either we accept that most of the probability scale ends up expressing doxastic neutrality—and that central applications of the theory have unattractive features that overshadow their usefulness—or a different translation manual is tolerated.

In light of these difficulties, the approach that I will take is to explore the perspectives that open up once the latter option is endorsed. But we must not forget that it comes with the prize of being unable to translate the sum of probabilities into the minimum of ranks as noted in section 2.1. This leaves us one step further away from a theoretical unification of ranking functions with probabilities, and it implies that we are only dealing with an approximation.

However, this does not mean that the translation would have been perfect on the infinitesimal alternative. In fact, Spohn (2012: 204) already has a list of 12 deviations. So if we are in any case looking at a less than perfect translation, and the infinitesimal interpretation is beset with the severe problems outlined in section 2.2 and 2.3, then we should be open for exploring ways of extending ranking theory that brings it into more contact with the psychological literature. Indeed, Spohn's own attitude towards the infinitesimal translation is that it is a nice formal possibility of unification to which we should not attribute any epistemological significance, however (2012: 205).

Yet, taking the line suggested above introduces another problem: once a different translation manual is accepted with a logarithmic base of $a \in (0,1)$, where $a \neq \varepsilon$, there is no *a priori* way of selecting a non-arbitrary value for a from the infinity of possible values. On the face of it, this realization is devastating to any attempt of deriving precise predictions from ranking theory, if it implies that any unsuccessful prediction could just be excused by claiming that the wrong logarithmic base had been chosen. This might seem to be a case of radical underdetermination by the empirical evidence. However, such an appearance would be misleading and it turns out that there is a pragmatic solution to this problem as we shall see in section 3.3.

3. Extending Ranking Theory by Logistic Regression

The purpose of section 3 is to extend the ranking theoretic approach to conditionals by logistic regression to enhance the former's use for experimental psychology. In so doing, the present model follows an old, venerable tradition in cognitive psychology of using an analogy between statistics and cognition in formulating new theories (see Gigerenzer & Murray, 1987 and Gigerenzer, 1988 for a detailed discussion).

As we shall see, the model to be introduced has the following nice qualities: (1) it provides equations that can produce quantitative predictions for the conditional inference task with only three parameters that are qualitatively constrained, (2) it throws new light on what is involved in performing the Ramsey test, (3) it allows us to introduce a numerical/verbal scale for two-sided ranking functions that has already found some

empirical support, and (4) it allows us to combine a theory of conditionals that is already well established in psychology of reasoning with accounts emphasizing probability raising.

3.1 Logistic Regression

In this subsection I will briefly introduce logistic regression, where we already begin to treat regression as a model of the knowledge representation underlying conditional reasoning in line with our general analogy between statistics and cognition. Its rationale is often introduced by first considering the simpler case of multiple linear regression.

The basic principle behind multiple linear regression is to represent variance of a measured or observed dependent variable (Y) as a weighted, linear function of the variance of a set of independent variables $\{X_1, \ldots, X_n\}$, which are under experimental control and function as predictors, plus some random noise (E) and the intercept (b_0):

$$\hat{Y} = b_0 + b_1 \cdot X_1 + \ldots + b_n \cdot X_n + E \tag{14}$$

The regression weights $\{b_1,\ldots,b_n\}$ determine how much the estimated dependent variable ($\hat{Y}$) changes with a one-unit change to the indexed predictor when all other variables are held constant. These weights are estimated from the data and express how much the indexed predictor contributes to reducing the variance in the dependent variable, when optimization methods are used to fit the model to the data (Eid, Gollwitzer, & Schmitt, 2010: ch. 16, 18).

For our purposes, using multiple linear regression as a model is beset with the following problem. In standard applications of multiple linear regression, the independent variables would be interval scaled or higher, and one of its presuppositions is that the estimated dependent variable, $\hat{Y}$, is interval scaled or higher. So we are faced with the problem of what the variables in the regression equation correspond to?

The tempting answer is to invoke probabilities (absolute scale) or ranking functions (ratio scale),[57] since they are used to illuminate the semantics of conditionals. But that this would be a mistake is easily seen once one starts to think about how multiple linear regression actually works. The cognitive system is hardly trying to use variance in its

ranking function or probability distribution function of the predictors to predict the variance in its ranking function or probability distribution function of the consequent. The reason why this sounds odd is that it would get the intentionality the wrong way around: presumably what is being predicted is something objective, like the truth values of propositions, and not an aspect of the cognitive system's own doxastic states. But propositions are normally treated as binary variables and they thus fail to meet the requirements of multiple linear regression.

To solve this problem logistic regression can be invoked, which depicts the probability of a binary dependent variable taking a particular value as a non-linear function of the values the independent variables take, where the independent variables can be scaled on any scale we like:

$$P(Y = 1|X_1, \ldots, X_n) = \frac{z}{1+z} \qquad \text{for } z = e^{b_0+b_1 \cdot X_1+b_2 \cdot X_2+\ldots+b_n \cdot X_n} \tag{15}$$

For our purposes it is moreover pleasing to note that what is being estimated is *the conditional probability* that the dependent variable takes a particular value. For the model is thereby able to come into contact with approaches to conditionals that emphasize probability raising (Douven 2008, 2013, Spohn, 2013a), which focuses on the relationship between P(C | A) and P(C | $\overline{\text{A}}$), and the suppositional theory of conditionals (Evans & Over, 2004), which takes Adam's thesis as its point of departure (cf. chapter II).

To simplify the calculations, (15) can be transformed into (15a):

$$P(Y = 1|X_1, \ldots, X_n) = \frac{1}{1+\frac{1}{z}} \tag{15a}$$

Furthermore, it should be noted that due to the non-additive and non-linear relationship between the independent variables and probabilities in (15), the effect of one independent variable (X_i) varies with the values of the other independent variables and the predicted probabilities. For this reason the effect of X_i cannot be fully represented by a single coefficient and instead has to be evaluated at a particular value, or set of values, which renders its interpretation cumbersome (Pampel, 2000: ch. 2).

It may therefore be useful to note that two transformations for (15) exist, where the effect of X_i can be summarized by a single coefficient. When (15) is stated in terms of *conditional odds* (O_i),[58] e^{b_i} represents the factor by which the odds are multiplied, when X_i increases by one-unit and all other variables are held constant:

$$O_i = e^{b_0+b_1\cdot X_1+b_2\cdot X_2+\ldots+b_n\cdot X_n} = e^{b_0}\cdot e^{b_1\cdot X_1}\cdot e^{b_2\cdot X_2}\cdot \ldots \cdot e^{b_n\cdot X_n} \qquad \text{for } i = 1,\ldots,n \qquad (15b)$$

And when (2) is stated in terms of *logged odds*, b_i represents how much $\hat{Y}$ changes with a one-unit change to the indexed predictor, when all other variables are held constant:

$$\ln(O_i) = b_0 + b_1\cdot X_1 + b_2\cdot X_2 + \ldots + b_n\cdot X_n \qquad \text{for } i = 1,\ldots,n \qquad (15c)$$

(15c) thus parallels the case of multiple linear regression (however the units have changed to logged odds).

3.2 Logistic Regression and Ranking Theory

At first glimpse it may seem puzzling what this statistical model has to do with ranking theory. But the relationship between the two will gradually unfold throughout this chapter. The purpose of this subsection is to introduce some initial observations.

As Spohn (2012: 76) points out, although using two-sided ranking functions may be the most intuitive way of presenting ranking theory, there is no simple axiomatization of them. Furthermore, since two-sided ranking functions appear to be a derived notion that is ultimately to be defined in terms of negative ranking functions, he prefers the latter as an epistemological tool. However, as we will begin to see, the former is attractive for psychological purposes. Moreover, as we will now see, two-sided ranking functions are not merely the derived notion that they appeared to be. In fact, they have their own interpretation.

Since logistic regression deals with *logits*, or *logged odds*, it is interesting to note that two-sided ranking functions give us a comparable metrics. However, the logarithmic bases differ and two-sided ranking functions are actually the logged odds of a proposition *not* taking the value 'true':

$$\tau(A) = \beta(A) - \beta(\overline{A}) = \kappa(\overline{A}) - \kappa(A)$$

$$\kappa(A) \approx \log_a\big(P(X=1)\big), \qquad \text{where } 0 < a < 1$$

$$\tau(A) \approx \log_a\big(P(X=0)\big) - \log_a\big(P(X=1)\big) = \log_a\left(\frac{P(X=0)}{P(X=1)}\right)$$

To understand why two-sided ranking functions take this form, it is useful to consider that:

$$\log_a\left(\frac{P(x)}{1-P(x)}\right) = -\log_a\left(\frac{1-P(x)}{P(x)}\right)$$

$$\log_{\frac{1}{a}}\left(\frac{1-P(x)}{P(x)}\right) = -\log_a\left(\frac{1-P(x)}{P(x)}\right)$$

Hence,

$$\log_{\frac{1}{a}}\left(\frac{1-P(x)}{P(x)}\right) = \log_a\left(\frac{P(x)}{1-P(x)}\right)$$

So when two-sided ranking functions are the logged odds of a proposition *not* taking the value 'true' with a logarithmic base of $a \in (0,1)$, they can always be rewritten as the logged odds of a proposition taking the value 'true' with the logarithmic base of a^{-1}.

Hence, if a logarithmic base of e^{-1} is chosen for ranking functions, it is possible to bring the two formalisms into contact, because the logarithmic base of our regression equations is e. As we shall see, this observation will later prove to be crucial, when we begin deriving predictions from a model based on logistic regression for conditional reasoning in the next section.

3.3 The Conditional Inference Task

When it comes to producing predictions for psychology of reasoning, it is important to consider existing experimental paradigms, because most of psychology of reasoning is organized around a few experimental paradigms that have been studied extensively (Manktelow, 2012). We will therefore continue our investigation of the parallel between logistic regression and ranking theory by focusing on a particular experimental paradigm.

In the conditional inference task the participants are asked to rate the conclusions of the following four inferences: MP (*modus ponens*: $p \rightarrow q$, $p \therefore q$), MT (*modus tollens*: $p \rightarrow q$, $\neg q \therefore \neg p$), AC (*affirmation of the consequent*: $p \rightarrow q$, $q \therefore p$), and DA (*denial of the antecedent*: $p \rightarrow q$, $\neg p \therefore \neg q$). Of these, only MP and MT are classically valid, if '$\rightarrow$' is read as the material implication. While MP is consistently endorsed nearly to the maximum degree (89-100% with abstract material), the finding that the logically valid MT is typically endorsed at only about 40-80%, and that the logically invalid AC and DA are typically endorsed with about 20-75%, is one of the key findings that have contributed to the current rationality debates in cognitive psychology about how appropriate deductive logic is as a normative model of human reasoning (Evans & Over, 2004: 46, Oaksford & Chater, 2007, Manktelow, 2012).

However, it is far from obvious that AC and DA should be seen as flaws of reasoning. After all, AC characterizes the type of abductive inference embodied in Bayes' theorem where we reason from an effect back to its potential cause,[59] which is characteristic of scientific reasoning. Bayes' theorem expresses this type of reasoning by requiring that we update our degree of belief in a hypothesis after the confirmation of one of its predictions, as is easily seen once it is expressed in the following form:

$$P(H \mid E) = \frac{P(E \mid H) \cdot P(H)}{P(E)} \leftrightarrow P(A \mid C) = \frac{P(C \mid A) \cdot P(A)}{P(C)}$$

Moreover, DA also has its justification in argumentative contexts when it is used to challenge a reason that has been offered in support of C thus urging that C has been advanced on an insufficient basis, as Godden & Walton (2004) argue.

It would therefore be wrong to dismiss the endorsement of these types of inferences as a symptom of irrationality merely because such inferences are connected with uncertainty and thereby fail to be validated by classical logic. Hence, it is an attractive

feature of Spohn's (2013a) relevance approach that it is not forced to render these inferences invalid. In fact, Spohn's (2013a) theory validates all four inferences (MP, MT, AC, and DA). In contrast, the suppositional theory of conditionals follows the horseshoe analysis in rendering AC and DA invalid (Evans & Over, 2004: 45).

Moreover, the ranking theoretic approach to conditionals is compatible with the asymmetry in the endorsement rates that has been found. Yet, it is unable to deliver any precise, quantitative predictions about these endorsement rates. This, however, is accomplished by the extension of the theory to be presented below.

To back up a little, what leads to the acceptance of AC and DA on the basis of Spohn's theory is that positive relevance is a symmetric relation:

$$\text{If } \tau(C \mid A) > \tau(C \mid \overline{A}), \text{ then } \tau(A \mid C) > \tau(A \mid \overline{C}) \tag{16}$$

So if A is positively relevant for C, then C is positively relevant for A. Moreover, as Spohn (2013a: 1092) points out, it also holds that:

$$\text{If A is positively relevant for C, } \sim\text{A is positively relevant for } \sim\text{C} \tag{17}$$

Which together with (16) yields contraposition:

$$\text{If A is positively relevant for C, } \sim\text{C is positively relevant for } \sim\text{A} \tag{18}$$

(16) validates AC, (17) validates DA, and (18) validates MT.

Finally, as Spohn (2013a) points out, these symmetrical relevance relations make room for explaining the varying endorsement rates for these four inferences, because although the relations run in both directions they need not do so to the same degrees.

However, this only provides us with a rough qualitative prediction of the results of the experiments on the conditional inference task. But it is definitely on the right track, insofar as a typical finding using abstract content and instructions stressing logical necessity is that MP > MT > AC ≥ DA, whereas the endorsement rates depend on perceived

sufficiency and necessity of the antecedent for the consequent, when it comes to realistic content in the absence of a conditional rule (Klauer, Beller, & Hütter, 2010).

What these ratings show is that we are not looking for a relation governed by perfect symmetry, when modeling the relationship between the antecedent and consequent in conditionals, because then we would end up with the bi-conditional interpretation, where MP, MT, AC, and DA should all be fully endorsed to the same degrees. On the other hand, the data don't support the material implication interpretation, whereby MP and MT should be fully endorsed while AC and DA should be fully rejected (Evans & Over, 2004). Instead what we see is that all four inferences are endorsed, but to different degrees, which requires a relationship between the antecedent and consequent that holds in both directions but to different degrees.

If we are to turn these observations into quantitative predictions, we can exploit the fact that something similar holds for logistic regression. First, it is useful to note that the following fact about linear regression has a counterpart in logistic regression. Correlation and linear regression are sometimes[60] distinguished by pointing out that the former is symmetric, whereas the latter is asymmetric in the following sense: in the case of correlation, no distinction is made between dependent and independent variables, whereas it makes a difference, which variables are treated as dependent and independent in a regression equation.

To be sure, it is possible to treat Y as a predictor of X instead of treating X as a predictor of Y by using table 5, where 's_Y' is the standard deviation, 's_{XY}' is the sample covariance, 'r_{XY}' is the sample correlation coefficient, and '$\bar{x}$' is the sample mean:

Table 5, Linear Regression

	X as a predictor of Y:	*Y as a predictor of X:*
Slope:	$b_1 = r_{XY} \cdot \frac{s_Y}{s_X} = \frac{s_{XY}}{s_X^2}$	$b_1^* = r_{XY} \cdot \frac{s_X}{s_Y} = \frac{s_{XY}}{s_Y^2}$
Intercept:	$b_0 = \bar{y} - b_1 \cdot \bar{x}$	$b_0^* = \bar{x} - b_1^* \cdot \bar{y}$

But the regression lines to which the scatter plot will be fitted will differ depending on whether X is treated as a predictor of Y or Y is treated as a predictor of X. It turns out

that something similar holds for logistic regression, when the independent variable is also a binary variable.[61]

With this in mind, we now turn to the asymmetry between when X is used as a predictor of Y and Y is used as a predictor of X in logistic regression as exhibited in table 6. As we notice, the slopes are identical,[62,63] but the intercepts differ:

Table 6, Logistic Regression

	X as a predictor of Y:	*Y as a predictor of X:*
Intercept:	$e^{b_0} = \frac{P(Y=1 \mid X=0)}{P(Y=0 \mid X=0)}$	$e^{b_0^*} = \frac{P(X=1 \mid Y=0)}{P(X=0 \mid Y=0)}$
"Slope":	$e^{b1} = \frac{P(Y=1, X=1)}{P(Y=0, X=1)} \cdot \frac{P(Y=0, X=0)}{P(Y=1, X=0)}$	

Accordingly, we have now reached a point, where we are able to see that the logistic regression equations give us a model of a predictor relationship that has the desired property of a relation that holds in both directions but to different degrees, which we observed above would be useful in modeling the endorsement rates of MP, MT, AC, and DA. Exploiting this fact, the following equations can be formulated for the reduced conditional inference problems in Klauer *et al.* (2010), where: (i) the participants are presented with the minor premise and conclusion of MP, MT, AC, and DA without the conditional rule, and they are accordingly being asked to rate the conclusion based on the minor premise alone (i.e. MP_R: p ∴ q. MT_R: ¬q ∴ ¬p. AC_R: q ∴ p. DA_R: ¬p ∴ ¬q), and (ii) the consequent (C = {Y = 1}, non-C = {Y = 0}) is also used as a predictor of the antecedent (A = {X = 1} and non-A = {X = 0}):

(MP_R) $$P(Y=1 \mid X=1) = \frac{1}{1+e^{-(b_0+b_1)}} \tag{19}$$

(AC_R) $$P(X=1 \mid Y=1) = \frac{1}{1+e^{-(b_0^*+b_1)}} \tag{20}$$

(DA_R) $$P(Y=0 \mid X=0) = \frac{1}{1+e^{b_0}} \tag{21}$$

(MT_R) $$P(X=0 \mid Y=0) = \frac{1}{1+e^{b_0^*}} \tag{22}$$

As we shall see later, these equations have a range of nice predictions. In section IV 2.2.2, the next step of modeling the presence of the conditional rule in MP, MT, AC, and DA will moreover be undertaken.

In section 3.2 we already noticed the close relationship between logistic regression, which has a logged odds format, and two-sided ranking functions, when a logarithmic base of e^{-1} is chosen. It is now possible to make the parallel even closer by considering (19)-(22) under a different light. In their logged odds format they take the following form:

$$(\mathrm{MP_R}) \qquad \ln\left(\frac{P(Y=1\mid X=1)}{P(Y=0\mid X=1)}\right)=b_0+b_1 \tag{23}$$

$$(\mathrm{AC_R}) \qquad \ln\left(\frac{P(X=1\mid Y=1)}{P(X=0\mid Y=1)}\right)=b_0^*+b_1 \tag{24}$$

$$(\mathrm{DA_R}) \qquad \ln\left(\frac{P(Y=0\mid X=0)}{P(Y=1\mid X=0)}\right)=-b_0 \tag{25}$$

$$(\mathrm{MT_R}) \qquad \ln\left(\frac{P(X=0\mid Y=0)}{P(X=1\mid Y=0)}\right)=-b_0^* \tag{26}$$

However, since the following holds:

$$\ln\left(\frac{P(Y=1\mid X=1)}{P(Y=0\mid X=1)}\right)=\log_{\frac{1}{e}}\left(\frac{P(Y=0\mid X=1)}{P(Y=1\mid X=1)}\right)=\tau(C\mid A)$$

$$\ln\left(\frac{P(X=1\mid Y=1)}{P(X=0\mid Y=1)}\right)=\log_{\frac{1}{e}}\left(\frac{P(X=0\mid Y=1)}{P(X=1\mid Y=1)}\right)=\tau(A\mid C)$$

$$\ln\left(\frac{P(Y=0\mid X=0)}{P(Y=1\mid X=0)}\right)=\log_{\frac{1}{e}}\left(\frac{P(Y=1\mid X=0)}{P(Y=0\mid X=0)}\right)=\tau(\overline{C}\mid\overline{A})$$

$$\ln\left(\frac{P(X=0\mid Y=0)}{P(X=1\mid Y=0)}\right)=\log_{\frac{1}{e}}\left(\frac{P(X=1\mid Y=0)}{P(X=0\mid Y=0)}\right)=\tau(\overline{A}\mid\overline{C})$$

We now see that:

$$(\mathrm{MP_R}) \qquad \tau(C\mid A)=b_0+b_1 \tag{27}$$

$$(\mathrm{AC_R}) \qquad \tau(A\mid C)=b_0^*+b_1 \tag{28}$$

(DA$_R$) $\tau(\overline{C} \mid \overline{A}) = -b_0$ (29)

(MT$_R$) $\tau(\overline{A} \mid \overline{C}) = -b_0^*$ (30)

And that table 6 can be reformulated on the basis of two-sided ranking functions as shown in table 7:

Table 7, Translation of Table 6 into Ranking Functions

	X as a predictor of Y:	*Y as a predictor of X:*
Intercept:	$b_0 = \tau(C \mid \overline{A})$	$b_0^* = \tau(A \mid \overline{C})$
"Slope":	$b_1 = \tau(C \mid A) - \tau(C \mid \overline{A}) = \tau(A \mid C) - \tau(A \mid \overline{C})$	

Table 7 makes the parametrization much more perspicuous than table 6 managed to. In the case of b_0 and b_0^*, we are dealing with a measure of our belief in the consequent, when the predictor takes the value 'false', whereas the b_1 parameter quantifies the *relevance* of the predictor for the consequent. We moreover observe that in spite of the fact that the absolute magnitudes of $\tau(C \mid A)$ and $\tau(C \mid \overline{A})$ may diverge from the magnitudes of $\tau(A \mid C)$ and $\tau(A \mid \overline{C})$ respectively, the differences in these pairs stay identical, and so the b_1 parameter stays the same no matter from which direction we view the predictor relationship.

To explain all the parallels we are observing between logistic regression and two-sided ranking functions it suffices to note that:

$$b_0 + b_1 = ln\left(\frac{P(Y=1|X=0)}{P(Y=0|X=0)}\right) + ln\left(\frac{\frac{P(Y=1|X=1)}{P(Y=0|X=1)}}{\frac{P(Y=1|X=0)}{P(Y=0|X=0)}}\right) = ln\left(\frac{P(Y=1|X=1)}{P(Y=0|X=1)}\right)$$

But, of course:

$$\ln\left(\frac{P(Y=1|X=1)}{P(Y=0|X=1)}\right) = log_{\frac{1}{e}}\left(\frac{P(Y=0|X=1)}{P(Y=1|X=1)}\right) = \tau(C|A)$$

And something similar holds for (28)-(30). In other words, it turns out that (19)-(22) can be derived from probabilistic transformations of two-sided ranking functions once a

logarithmic base of e^{-1} is chosen. This observation is extremely useful, because it implies that we can use (19)-(22) to derive precise quantitative predictions for what had to remain qualitative predictions in Spohn (2013a). In section 3.5 we will see exactly how rich these predictions turn out to be.

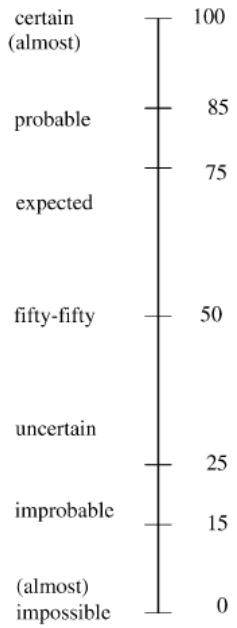
At this point it is only appropriate that we return to the issue raised in section 2.4 about the arbitrariness of selecting a logarithmic base for ranking functions and the worry that it will have the implication that the theory will end up being radically, empirically underdetermined once the infinitesimal translation manual has been rejected, because our model makes use of a logarithmic base of e^{-1}.

The first thing to notice is although there is no *a priori* basis for selecting a logarithmic base other than the infinitesimal base, this doesn't mean that we are completely without constraints. In particular, we saw that the main problem with the infinitesimal translation was that it seemed to fit too poorly with the way humans carve up the probability scale. This suggests that our choice of a logarithmic base should be constrained empirically. In this context, it is worth noticing that Spohn (2013a) suggests that it would be possible to align ranking functions with the linguistic qualifiers we use to express our degrees of beliefs. This suggests that independent evidence of the numerical values that ordinary participants associate with verbal expressions of degrees of beliefs should be used in selecting the logarithmic base.

If a logarithmic base of e^{-1} is chosen, it will be possible for the ranks to spread out more widely over the probability scale, which gives us the following scale:

$\tau(\cdot \mid \cdot)$	-3	-2	-1	0	1	2	3
$P(\cdot \mid \cdot)$	$\frac{1}{1+e^3}$	$\frac{1}{1+e^2}$	$\frac{1}{1+e}$	0.5	$\frac{e}{1+e}$	$\frac{e^2}{1+e^2}$	$\frac{e^3}{1+e^3}$
	0.05	0.12	0.27		0.73	0.88	0.95

Incidentally, this scale fits nicely with the following scale, which has already received empirical support (e.g. Witteman & Renooij, 2002) and been successfully used for eliciting expert knowledge for Bayesian networks (van der Gaag, Renooij, Schijf, Elbers, Loeffen, 2012):



However, it is possible that this scale may eventually be replaced by other scales that are better able to capture the linguistic phenomenology of expressing degrees of beliefs. So the policy that I will adopt is continue to use a logarithmic base of e^{-1} for illustrative purposes and be prepared to revise the equations, if another grading receives *independent support*. To the extent that such evidence is independent of the performance of the model on the conditional inference task, its calibration by it should not be seen as a question begging attempt to dodge unpleasant challenges.

The second thing to note is that as far as model fitting goes, it actually doesn't matter exactly which logarithmic base we select. The reason is that (19)-(22) have three parameters that will have to be estimated on the basis of the data. So if the logarithmic base is changed, the effect will just be to change the order of the magnitude of the estimated regression weights. So the problem of the lack of a principled basis for choosing a logarithmic base will not prevent its use for experimental purposes.

The third thing to note is that the *a priori* predictions that will be derived in section 3.5 apply to most values of the logarithmic base within the interval (0,1).

So the upshot is that section 3.5 delivers a set of predictions that can be used with (almost) any logarithmic base within this interval, and any use of the model that goes beyond this will rely on fitting the model's free parameters to the data, where a change of the logarithmic base merely has the effect of changing the order of the magnitude of the estimated parameters. The only difference that this will make is, however, to change the conventions for interpreting the size of the estimated coefficients.

3.4 Introducing Qualitative Constraints on the Free Parameters

As said, the model provided in section 3.3 has three parameters for each set of inferences. We will now see how one can introduce qualitative constraints on the values assigned to the estimated parameters. In order to do so, it is useful to keep the theoretical background in mind. In psychology of reasoning there has been a focus on the influence of *disablers* and *alternative antecedents* on conditional reasoning, which has *inter alia* been used to measure the influence of content on deductive reasoning. Disablers are conditions that prevent the consequent from obtaining even when the antecedent obtains and alternative antecedents are conditions other than the antecedent that are sufficient for the obtaining of the consequent. So if we take the conditional 'if the key is turned, the car will start', 'the battery is dead' would be a disabler and 'the car has been hot-wired' would be an alternative antecedent.

There is an experimental paradigm going back to Cummins (1995), which has studied how the endorsement rates of MP, MT, AC, and DA in causal inferences are affected by changes in the perceived sufficiency and necessity of the antecedent for the consequent, which has been induced by manipulating the availability of disablers and alternative antecedents. The general finding is that endorsement rates of AC and DA decrease with the availability of alternative antecedents and the endorsement rates of MP and MT decrease with availability of disablers (Politzer & Bonnefon, 2006).

Subsequent models of conditional reasoning have focused on integrating a component, which takes activation of memory traces of disablers and alternative antecedents into account (de Neys 2010, Cummins 2010). Moreover, studies based on means-end relations, permission, precaution, promises, tips, warnings, threats, temporal

relations, and obligations have shown that the phenomenon generalizes beyond causal inferences (Politzer & Bonnefon, 2006: 497, Verbrugge, Dieussaert, & Schaeken, 2007, Beller, 2008, see also Oaksford & Chater, 2010b).

Due to this theoretical background, it is a nice feature of our model that it is able to integrate the influence of disablers and alternative antecedents. Indeed, the illumination that the present account brings to this issue goes beyond this, because through Spohn's (2012: ch. 6) notion of sufficient and necessary reasons, we are able to make sense of the talk in the psychological literature about *degrees* of perceived sufficiency and necessity by pointing out that the former can be cashed out in terms of how far above 0 $\tau(C \mid A)$ is and the latter can be cashed out in terms of how far below 0 $\tau(C \mid \overline{A})$ is.[64]

The way the model integrates the influence of disablers and alternative antecedents is by the intended interpretation of its parameters as outlined in table 6. According to it, disablers should have the influence of *decreasing* the b_1 parameter and *increasing* the b_0* parameter (in virtue of increasing P(Y=0, X=1)), and alternative antecedents should have the influence of *decreasing* the b_1 parameter and *increasing* the b_0 parameter (in virtue of increasing P(Y=1, X=0)). This introduces a qualitative constraint on the values assigned to the parameters of the model.[65] And when it comes to model fitting, it is of advantage to have a model that is constrained by its intended interpretation as opposed to merely being able to model the data through the flexibility generated by its free parameters.

3.5 Deriving Predictions from the Logistic Regression Model

In deriving predictions from our model, it is useful to return to the specifications of the notions of reason and relevance from section 2.3, extend it to cover all the cases of when A is a reason against C, and substituting regression weights for two-sided ranking functions:

(31)	**A is positively relevant to C iff**	$\mathbf{b_0 + b_1 > b_0}$
	A is a supererogatory reason for C iff	$b_0 + b_1 > b_0 > 0$
	A is a sufficient reason for C iff	$b_0 + b_1 > 0 \geq b_0$
	A is a necessary reason for C iff	$b_0 + b_1 \geq 0 > b_0$
	A is an insufficient reason for C iff	$0 > b_0 + b_1 > b_0$

(32)	**A is irrelevant to C iff**	$\mathbf{b_0 + b_1 = b_0}$
(33)	**A is negatively relevant to C iff**	$\mathbf{b_0 > b_0 + b_1}$
	A is a supererogatory reason against C iff	$0 > b_0 > b_0 + b_1$
	A is a sufficient reason against C iff	$b_0 \geq 0 > b_0 + b_1$
	A is a necessary reason against C iff	$b_0 > 0 \geq b_0 + b_1$
	A is an insufficient reason against C iff	$b_0 > b_0 + b_1 > 0$

As we notice, A is *positively relevant* for C whenever $b_1 > 0$, A is *irrelevant* to C whenever $b_1 = 0$, and A is *negatively relevant* for C whenever $b_1 < 0$. Using this observation and the inequalities in (31)-(33), it is possible to derive predictions about the endorsement rates of MP_R, MT_R, AC_R, and DA_R for different types of reason relations.

As said, the beauty of these predictions is that they apply to most values of logarithmic bases within the interval of (0,1). But for illustrative purposes, we will continue to use a logarithmic base of e^{-1}.

A is a Sufficient Reason for C:

$$b_0 + b_1 > 0 \leftrightarrow \frac{1}{1+e^{-(b_0+b_1)}} > \frac{1}{2} \leftrightarrow MP_R > \frac{1}{2}$$

$$b_0 \leq 0 \leftrightarrow \frac{1}{1+e^{b_0}} \geq \frac{1}{2} \leftrightarrow DA_R \geq \frac{1}{2}$$

As we have already noted, the degree of sufficiency can be experimentally manipulated through disablers, which in turn increase P(Y=0, X=1). The absence of disablers should thus have the effect of increasing b_1 and decreasing b_0^* (cf. table 6). As a result, the absence of disablers should have the effect of increasing the endorsement of MP_R:

$$\frac{1}{1+e^{-(b_0+b_1)}} < \frac{1}{1+e^{-(b_0+b_1+a)}} \text{ for } a > 0$$

And MT_R:[66]

$$\frac{1}{1+e^{b_0^*}} < \frac{1}{1+e^{b_0^*-a}} \quad \text{for } a > 0$$

A is a Necessary Reason for C:

$$b_0 + b_1 \geq 0 \leftrightarrow \frac{1}{1+e^{-(b_0+b_1)}} \geq \frac{1}{2} \leftrightarrow MP_R \geq \frac{1}{2}$$

$$b_0 < 0 \leftrightarrow \frac{1}{1+e^{b_0}} > \frac{1}{2} \leftrightarrow DA_R > \frac{1}{2}$$

As we have already noted, the degree of necessity can be experimentally manipulated through alternative antecedents, which in turn affect P(Y=1, X=0). The presence of alternative antecedents should thus have the influence of decreasing b_1 and increasing b_0 (cf. table 6). As a result, the presence of alternative antecedents should have the effect of decreasing endorsements of DA_R:

$$\frac{1}{1+e^{b_0}} > \frac{1}{1+e^{b_0+a}} \qquad \text{for } a > 0$$

And AC_R:

$$\frac{1}{1+e^{-(b_0^*+b_1)}} > \frac{1}{1+e^{-(b_0^*+b_1-a)}} \qquad \text{for } a > 0$$

A is a Supererogatory Reason for C:

$$b_0 + b_1 > 0 \leftrightarrow \frac{1}{1+e^{-(b_0+b_1)}} > \frac{1}{2} \leftrightarrow MP_R > \frac{1}{2}$$

$$b_0 > 0 \leftrightarrow \frac{1}{1+e^{b_0}} < \frac{1}{2} \leftrightarrow DA_R < \frac{1}{2}$$

In contrast to sufficient and necessary reasons, supererogatory and insufficient reasons are not conceptually distinguished in the experimental literature. To experimentally manipulate supererogatory reasons, not only the presence of alternative antecedents should be manipulated but also their obtainance. So whereas a necessary reason would require the absence of alternative antecedents, and thus require high ratings of AC_R and DA_R, a supererogatory reason would require the presence and obtainance of alternative antecedents and thus require low ratings for AC_R and DA_R. Necessary and supererogatory

reasons can moreover be distinguished by the prediction displayed above that whereas the probability of DA_R should be > 0.5 and MP_R should be ≥ 0.5 for the former, DA_R should be < 0.5 and MP_R should be > 0.5 for the latter.

A is an Insufficient Reason for C:

$$b_0 + b_1 < 0 \leftrightarrow \frac{1}{1+e^{-(b_0+b_1)}} < \frac{1}{2} \leftrightarrow MP_R < \frac{1}{2}$$

$$b_0 < 0 \leftrightarrow \frac{1}{1+e^{b_0}} > \frac{1}{2} \leftrightarrow DA_R > \frac{1}{2}$$

To experimentally manipulate insufficient reasons, not only the presence of disablers should be manipulated but also their obtainance. So whereas a sufficient reason would require the absence of disablers, and thus require high ratings of MP_R and MT_R, an insufficient reason would require low ratings of MP_R and MT_R (which is known as *the suppression effect* in the psychological literature).[67] Sufficient and insufficient reasons can moreover be distinguished by the prediction that whereas MP_R should be > 0.5 and DA_R should be ≥ 0.5 for the former, MP_R should be < 0.5 and DA_R should be > 0.5 for the latter.

A is Irrelevant for C:

Since the prevailing theories in psychology of reasoning don't take the dimension of relevance into account, all of the predictions of patterns of conditional reasoning under manipulations of relevance hold the prospect of being *unique* to the present model.

In the case of irrelevance, conditionalizing on the antecedent will not affect the probability of the consequent. It thus holds that:

$$e^{b_0^*} = \frac{P(X=1|Y=0)}{P(X=0|Y=0)} = \frac{P(X=1)}{P(X=0)} \leftrightarrow b_0^* = ln\left(\frac{P(X=1)}{P(X=0)}\right)$$

$$e^{b_0} = \frac{P(Y=1|X=0)}{P(Y=0|X=0)} = \frac{P(Y=1)}{P(Y=0)} \leftrightarrow b_0 = ln\left(\frac{P(Y=1)}{P(Y=0)}\right)$$

This observation together with our earlier observation that $b_1 = 0$ for irrelevance can be used to derive predictions for content manipulations of the prior probability of the antecedent and the consequent:

Table 8, Predictions for Irrelevance Cases

	P(C) > 0.5	*P(C) = 0.5*	*P(C) < 0.5*
P(A) > 0.5	$AC_R > MT_R$, $MP_R > DA_R$	$AC_R > MT_R$, $MP_R = DA_R$	$AC_R > MT_R$, $MP_R < DA_R$
P(A) = 0.5	$AC_R = MT_R$, $MP_R > DA_R$	$AC_R = MT_R$, $MP_R = DA_R$	$AC_R = MT_R$, $MP_R < DA_R$
P(A) < 0.5	$AC_R < MT_R$, $MP_R > DA_R$	$AC_R < MT_R$, $MP_R = DA_R$	$AC_R < MT_R$, $MP_R < DA_R$

The following example illustrates the approach for P(A) > 0.5, P(C) > 0.5:

$$P(X = 1) > 0.5 \leftrightarrow b_0^* > 0 \leftrightarrow \frac{1}{1+e^{-b_0^*}} > \frac{1}{1+e^{b_0^*}} \leftrightarrow AC_R > MT_R$$

$$P(Y = 1) > 0.5 \leftrightarrow b_0 > 0 \leftrightarrow \frac{1}{1+e^{-b_0}} > \frac{1}{1+e^{b_0}} \leftrightarrow MP_R > DA_R$$

What the predictions in table 8 show is that when the antecedent is irrelevant for (or statistically independent of) the consequent, the model predicts that the four inferences coincide with what one would arrive at by using the prior probability of the conclusion while ignoring the probability of the premise. Intuitively, this seems exactly right.

Moreover, it holds in general that:

Table 9, Further Predictions for the Irrelevance Case

P(A) > P(C)	$MT_R < DA_R$, $AC_R > MP_R$
P(A) = P(C)	$MT_R = DA_R$, $MP_R = AC_R$
P(A) < P(C)	$MT_R > DA_R$, $AC_R < MP_R$

To illustrate, if P(A) = P(C) then:

$$\frac{1}{1+e^{b_0^*}} = \frac{1}{1+e^{b_0}} \leftrightarrow MT_R = DA_R$$

$$\frac{1}{1+e^{-b_0^*}} = \frac{1}{1+e^{-b_0}} \leftrightarrow AC_R = MP_R$$

Again the predictions in table 9 are also what one would expect for cases, where the antecedent is irrelevant for the consequent, insofar as the probabilities of the conclusions coincide with their prior probabilities.

Finally, it can be observed that $MP_R = 1 - DA_R$ and $AC_R = 1 - MT_R$ as $b_1 = 0$ in the case of irrelevance:

$$\frac{1}{1+e^{-b_0}} = 1 - \frac{1}{1+e^{b_0}}$$

$$\frac{1}{1+e^{-b_0^*}} = 1 - \frac{1}{1+e^{b_0^*}}$$

A is a Reason against C:

One way to view cases, where the antecedent is a reason *against* the consequent is to view them as negating the consequent of cases of positive relevance. As a result, if A is a sufficient reason *for* C, then A is *ipso facto* also a sufficient reason *against* ~C. To see that this is so, it is easiest to use the probabilistic version of (31) and (33):

$$P(Y = 1 \mid X = 1) > 0.5 \leftrightarrow P(Y = 0 \mid X = 1) < 0.5$$

$$0.5 \geq P(Y = 1 \mid X = 0) \leftrightarrow P(Y = 0 \mid X = 0) \geq 0.5$$

So:

$$P(Y = 1 \mid X = 1) > 0.5 \geq P(Y = 1 \mid X = 0) \leftrightarrow P(Y = 0 \mid X = 0) \geq 0.5 > P(Y = 0 \mid X = 1)$$

Similarly, it can be shown that if A is a supererogatory reason *for* C, then A is a supererogatory reason *against* ~C, if A is a necessary reason *for* C, then A is a necessary reason *against* ~C, and if A is an insufficient reason *for* C, then A is an insufficient reason *against* ~C. To emphasize this connection it may be useful to reformulate (33), so that it becomes perspicuous that if the relations in (31) hold for C, then the following holds for its negation:

(34) **A is negatively relevant to ¬C iff** **$-b_0 > -b_0 - b_1$**

A is a supererogatory reason against ¬C iff $0 > -b_0 > -b_0 - b_1$

A is a sufficient reason against ¬C iff $-b_0 \geq 0 > -b_0 - b_1$

A is a necessary reason against ¬C iff	$-b_0 > 0 \geq -b_0 - b_1$
A is an insufficient reason against ¬C iff	$-b_0 > -b_0 - b_1 > 0$

What this shows is that if we have a reason, A, *against* C that takes one of the four forms in (33), then the predictions specified for the corresponding positive relevance relation will hold for when A is taken as a reason *for* ~C and *vice versa*.

3.6 The Ramsey Test & the Suppositional Theory of Conditionals

After this inspection of some of the predictions that can be derived from the logistic regression model, it is time to return to the more theoretical side. By equating P(if *p*, *q*), or *acc*(if *p*, *q*), with P(*q*|*p*), the Ramsey test is used in the suppositional theory to estimate the probability of natural language conditionals (Evans & Over, 2004). What it requires is that the subject adds *p* to his knowledge base and estimates P(*q*) on this basis. Yet, how exactly this is carried out is not entirely clear as Over, Hadjichristidis, Evans, Handley, & Sloman (2007) point out:

> Explaining how the Ramsey test is actually implemented—by means of deduction, induction, heuristics, causal models, and other processes—is a major challenge, in our view, in the psychology of reasoning. (63)

That is, the Ramsey test does not explain how P(*q*) is determined once *p* has been added to the subject's knowledge base. Here the model allows us to come up with the following elegant suggestion: upon adding the antecedent to our belief set, its weight as a predictor of the consequent is used to compute the posterior probability.

Once this computational task has been formulated, it becomes possible to start theorizing about the cognitive processes carrying out the computations (e.g. fast and frugal heuristics) and mediating factors, which could influence this computation of the posterior probability in virtue of the regression weights. In particular, knowledge about causal models may influence the assigned weight and judgments of a hypothesis's virtues as an explanation may influence the weight in the case of use of Y as a predictor of X.

Moreover, the theoretical importance of this suggestion about the Ramsey test can be explicated in the following manner. According to Evans & Over (2004) and Evans (2007), 'if then' is a linguistic device that is used to simulate possibilities by activating a mental algorithm that makes us probe our background knowledge according to the Ramsey test.[68] However, although it makes a great deal of sense to say that simulating possibilities is useful from an evolutionary perspective, simulating possibilities is not by itself evolutionary useful, when the antecedent is irrelevant for the consequent. This suggests that the dimension of relevance adds to the idea of the function of the conditional as consisting in simulating possibilities. More generally, conditionals can be thought as serving an important communicative function in sharing knowledge about predictor relationships, which is seen in particular with indicative conditionals containing the predictive modal 'will' as in 'if it rains the match *will* be canceled' (cf. Dancygier, 1998, Dancygier & Sweetser, 2005). From this perspective, one of the main points of simulating possibilities can be seen as consisting in evaluating, whether the information on offer codifies useful information that the subject can adopt to improve his/her ways of coping with the uncertain environment. So when a speaker states an indicative conditional, the hearer can be seen as using 'if then' as a guide that possibilities are to be simulated, because the consequent is to be evaluated under the supposition of the antecedent (in agreement with Evans & Over (2004)). Yet, the evolutionary point of this exercise consists in its being a way of evaluating, whether useful predictive information is being shared. Accordingly, the hearer should view it as a failure, if the antecedent is irrelevant and leaves the probability of the consequent unchanged. We thus begin to see how relevance considerations may enter into this process of mental simulation in accordance with the suggestion of the computational task involved in performing the Ramsey test outlined above.

Indeed, it is possible to go further than this and establish a link to Rescorla and Wagner's work on classical conditioning by saying that the information shared by indicative conditionals containing the predictive modal 'will' is a linguistic counterpart of the kind of information acquired in classical conditioning. For the classification that Granger & Schlimmer (1986: 150) attribute to Rescorla in the following quote corresponds exactly to

the probabilistic version of Spohn's (2013a) analysis of positive relevance, negative relevance, and irrelevance:

> experiments explicitly aimed at exploring the space of possible contingencies led Rescorla to form the characterization that if $p(\text{US}|\text{CS}) > p(\text{US}|\overline{\text{CS}})$, then excitatory conditioning occurs, and if $p(\text{US}|\text{CS}) < p(\text{US}|\overline{\text{CS}})$, then inhibitory conditioning occurs, and if $p(\text{US}|\text{CS}) = p(\text{US}|\overline{\text{CS}})$, then neither type of conditioning occurs (Rescorla, 1966, 1967, 1968, 1969). ['US' = unconditioned stimuli. 'CS' = conditioned Stimuli, NSO]

Having thus introduced the comparison with the Ramsey test above, it becomes interesting to ponder more generally about the relationship between the logistic regression model and the suppositional theory of conditionals. One striking contrast is that the suppositional theory of conditionals, as it is elaborated in Edgington (1995), Bennett (2003), and Evans & Over (2004), attempts to account for the semantics of indicative conditionals on the basis of the Ramsey test and Adam's thesis without incorporating the relevance dimension as emphasized by Spohn (2013a), Douven (2008, 2013), and others. As a result, considerations of probability raising will at most be relegated to issues of pragmatics and the assertability of conditionals along with Gricean principles, which are routinely invoked in an *ad hoc* manner in the psychological literature, when the theories are facing their limits.

In chapter II, a general argument has already been presented against this tendency. However, the point that I now want to make is that this assumption might appear to be justified empirically to the extent that Over *et al.* (2007), and a recent unpublished experiment by Klauer and Singmann in Freiburg,[69] have failed to find evidence for relevance being a substantial predictor of the assigned probabilities of conditionals. Yet, the way the idea of probability raising was investigated was by testing whether: (a) P(if *p, q*) was given by the delta-p rule (i.e. whether P(if *p, q*) = P(*q*|*p*) - P(*q*|¬*p*)), and (b) P(*q*|¬*p*) would come out as a significant predictor in regression analyses of the mean ratings of P(if *p, q*). So another way of viewing these experiments is that perhaps this was just not a good way of implementing the idea of probability raising.

As we know from table 7, the b_1 parameter also expresses a delta value in conditional degrees of beliefs. But it is stated in terms of two-sided ranking functions rather than in terms of probabilities (i.e. $b_1 = \tau(C \mid A) - \tau(C \mid \overline{A}) = \tau(A \mid C) - \tau(A \mid \overline{C})$). When translated into probabilities, what it amounts to is actually not a difference in conditional probabilities, but the logarithm of an *odds ratio*:

$$b_1 = ln\left(\frac{\frac{P(Y=1|X=1)}{P(Y=0|X=1)}}{\frac{P(Y=1|X=0)}{P(Y=0|X=0)}}\right) = ln\left(\frac{\frac{P(X=1|Y=1)}{P(X=0|Y=1)}}{\frac{P(X=1|Y=0)}{P(X=0|Y=0)}}\right)$$

That is, this parameter captures the idea of probability raising by representing how much the logged odds of Y = 1 rises, when the antecedent takes the value '1' as opposed to taking the value '0'.

As it stands, the evidence supports the equation of P(if *p*, *q*) = P(*q* | *p*),[70] which seems to violate the idea of probability raising. But luckily we can have our cake and eat it, because as we have seen, the logistic regression model gives us a conditional probability as its output (in agreement with this equation). Yet, the antecedent still turns out to be probability raising with respect to Y = 1 as long as $b_1 > 0$. So (19)-(22) turn out to be a nice compromise between the idea of probability raising and the equation of P(if *p*, *q*) = P(*q* | *p*).[71] However, this does not mean that the present model does not supply any experimentally distinguishable predictions of its own. For as we have seen, even if (19)-(22) and the suppositional theory share a conditional probability as their output, its value and the patterns of conditional reasoning are influenced by the dimension of relevance through the b_1 parameter. As a result, the logistic regression model is able to predict a general effect of relevance in the probabilities assigned to conditionals and in patterns of conditional reasoning, which holds that there should be statistical significant differences between sets of conditionals that have been pretested to fall in the positive relevance, irrelevance, or negative relevance group.[72]

Appendix 2:
An Alternative Taxonomy of Reason Relations

The purpose of this appendix is to present some considerations that might make us prefer alternative taxonomies of reason relations to Spohn's, which we have been working with throughout this chapter.

To recapitulate, Spohn's taxonomy of reason relations takes the following form:

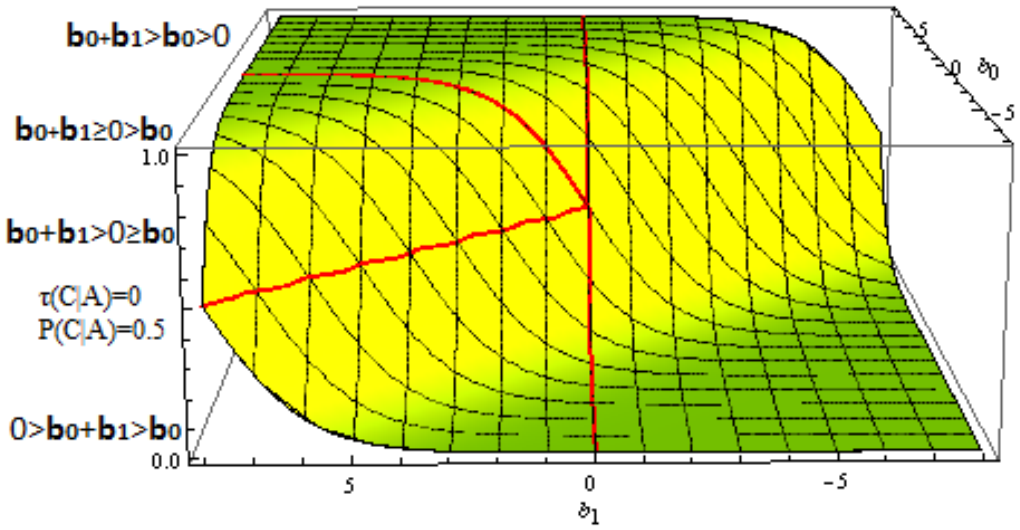


Figure 1: a visualization of Spohn's (2012: ch. 6) taxonomy of reason relations, $\tau(C \mid \overline{A}) = b_0$, $\tau(C \mid A) = b_0 + b_1$.

On this visualization, cases where A is an *insufficient* reason for C are located in the triangle in the lower left corner, cases where A is a *supererogatory* reason for C are located in the square in the upper left corner, and cases where A is a *sufficient* reason or a *necessary* reason are located in the triangle in between. Cases where A is a sufficient reason for C include the upper red edge of this triangle but don't include the lower red edge; cases where A is a necessary reason for C include the lower red edge of this triangle but don't include the upper red edge. As we see, according to this taxonomy sufficient and necessary

reasons are the only categories that are capable of applying simultaneously; all the other categories are mutually exclusive.

However, if we use the notion of supererogatory reasons to analyze cases of causal overdetermination as Spohn (2012: ch. 14) does, then we might be interested in allowing for instances of the former, where several, sufficient conditions are occurring simultaneously. As a result the following more generic notion of sufficient reasons might be preferable for some purposes. A is a sufficient reason for C iff: (i) $\tau(C \mid A) > \tau(C \mid \overline{A})$, and (ii) $\tau(C \mid A) > 0$.

Similarly, as philosophers are often interested in necessary reasons that are jointly sufficient but insufficient in themselves, it might be preferable to work with the following generic notion of a necessary reason for some purposes. A is a necessary reason for C iff: (iii) $\tau(C \mid A) > \tau(C \mid \overline{A})$, and (iv) $\tau(C \mid \overline{A}) < 0$.

As a result, cases where A is a sufficient reason for C will now be located above the red edge separating the two triangles and include the square in the upper left corner. Furthermore, cases where A is a necessary reason for C will now be located below the red edge separating the square and the middle triangle and include the triangle in the lower left corner. As we notice, it is still possible for A to be a sufficient and necessary reason for C as before, but the notions of sufficiency and necessity have been made more broad to allow for the cases identified above.

As Eric Raidl has pointed out, one further advantage of this taxonomy is that while Spohn's taxonomy is forced to hold that deductive reasons are *either* supererogatory *or* sufficient reasons (Spohn, 2012: 110), the present is able to hold the more plausible view that deductive reasons are always sufficient reasons. The problem is that while deductive reasons analyzed as the subset relation have the desirable property of not being relativized to epistemic states, as Spohn (ibid: 109) points out, whether deductive reasons end up being supererogatory or sufficient reasons on his analysis depends on whether the consequent is believed in advance. So relativity to epistemic states with respect to the status of deductive reasons is introduced through the backdoor on Spohn's taxonomy.

However, as it stands the generic notions introduced above are afflicted by the problem of always making Spohn's notion of supererogatory reasons sufficient reasons and

always making Spohn's notion of insufficient reasons necessary reasons. To remedy this defect, further conditions could be introduced in the case of sufficient reasons in addition to the ones already stated. In particular, it could be required for every B—such that B is a reason for C, and B is neither identical to A nor an enabling condition for A being a reason for C[73]—that A would still raise the two-sided rank of C above zero even if B were to take the value 'false'. Due to the monotonicity of deductive reasons, it would still hold that deductive reasons are always sufficient reasons on this expanded version.

The underlying intuition is that even if the prisoner had still been executed in case the third soldier hadn't made his shot—because there were 5 other soldiers firing simultaneously—the shot to the heart by the third soldier is a sufficient reason for believing in the death of the prisoner just in case that it alone could have killed the prisoner in the absence of the other shots. That is, if the shot to the heart was a sufficient reason to believe in the death of the prisoner in itself, it should remain a sufficient reason for believing in the death of the prisoner irrespectively of whether any further sufficient reasons obtain. For some of the other rank-raising, supererogatory reasons this condition will not be satisfied. For instance, if the shot of the fourth soldier hit some non-vital organs, then it may have speeded up the process of the death of the prisoner (so that we believe even more firmly that he is going to die when it is added to all the other shots), but where he could have survived it in the absence of the others. That is, although adding our third condition to condition (i) and (ii) allows for supererogatory reasons to be sufficient reasons, it no longer holds that supererogatory reasons automatically become sufficient reasons on our expanded notion.

Moreover, further conditions could be introduced for the notion of necessary reasons in addition to the ones already stated. In particular, it could be required for every D—such that D is a reason for C and D is not identical to A—that the falsity of A would lower the two-sided rank of C below zero even if D were to take the value 'true'.

Here the underlying intuition is that in order for A to be a necessary reason for C it must hold that C is disbelieved if A is false irrespectively of what else the agent believes. In this way we ensure that necessary reasons can either be sufficient or insufficient reasons

(depending on what happens when A is true), but insufficient reasons are not automatically necessary reasons, because some of them fail to satisfy our third condition.

---

[51] Acknowledgement: this chapter profited from discussions with Wolfgang Spohn, Karl Christoph Klauer, Sieghard Beller, Henrik Singmann, Eric Raidl, Igor Douven, and the audience of the third annual meeting of New Frameworks of Rationality.

[52] Potential Objection: as the reader will notice, there is at least one contender that is salient through its absence in chapter II; to wit, possible worlds semantics, which holds that a conditional is true if the consequent is true in the nearest possible world in which the antecedent is true. Mainly, this approach is used in the analysis of counterfactuals and our focus was on indicative conditionals, but Stalnaker has applied it to indicative conditionals as well. The reason for its absence is that although it gets mentioned in the psychological literature (e.g. Evans & Over, 2004), it is not really employed in experimental psychology due to the difficulty with deriving predictions from it.

[53] Extension: as Spohn (2012: ch. 5) points out, it is also possible to introduce a threshold notion, where **Bel**(A) iff $\kappa(\overline{A}) > z$ for $z > 0$ and beliefs would still be deductively closed. However, it requires that we can make sense conceptually of degrees of neutrality.

[54] Prospects of a direct application of ranking theory to psychology of reasoning: indeed, Gabriele Kern-Isberner and her group at Dortmund University plan to adopt this direct approach. Of course, in such an endeavor it becomes interesting to test the differences between degrees of beliefs expressed in probabilities and in ranking functions. For instance, the reduction in computational complexity involved in only having to deal with plus and minuses instead of with multiplications and divisions might be of psychological consequence. Interestingly, Juslin, Winman, & Lindskog (2011) present some first, preliminary results that it is possible to reduce the famous base-rate neglect by presenting problems in a logged probability format, where information only has to be linearly combined, instead of presenting them in a probabilistic format, where multiplications and divisions are required. Moreover, as the conjunction rule for positive ranks equates the degree of belief in the conjunction with the degree of belief in the least believed conjunct, it turns out that there should actually not be a difference whether the participants choose 'Linda is a bank teller' or 'Linda is a bank teller and a feminist' in the famous "conjunction fallacy", according to ranking theory. And as the second option of the two is more in agreement with the background knowledge introduced in the instruction, one could treat this conjunction rule as the contribution of the form mode of reasoning within a dual source approach (cf.

chapter IV), and argue that the content component should make it more rational to favor 'Linda is a bank teller and a feminist' in spite of the fact that this is normally taken to be the *incorrect* response.

[55] On translating negative ranks of zero into probabilities: in relation to the translation of rank zero into probabilities, the following issue emerges: theorem 10.1 in Spohn (2012: 203) says that if we take the standard part of the logarithm with an infinitesimal base of a probability, then we end up with a real valued negative rank $\left(\kappa(A) = st\left(\log_i[P(A)]\right)\right)$. However, as such it is silent on what happens, when we go in the other direction and have ranking functions that we want to translate into probabilities. But there is the difficulty that $\kappa(A) = 0 \leftrightarrow \log_a[P(A)] = 0 \leftrightarrow P(A) = a^0 \leftrightarrow P(A) = 1$. The reason is that such a translation is affected by the problem that the representation of doxastic neutrality as $\kappa(\overline{A}) = \kappa(A) = 0$ would violate the axioms of the probability calculus as the probabilities sum up to more than one. It is clear that a principled solution is needed for dealing with this problem. Spohn (personal communication) has suggested that the probabilities need to be redistributed such that they in any case sum up to 1. Accordingly, once all other ranks have been translated, the rank zeros receive an equal share of whatever is left.

[56] On how to express Spohn's taxonomy of reason relations into probabilities: it should be noted that treating $\tau(C \mid A) = 0$ as the point of doxastic indifference for two-sided ranking functions commits the theory to treating P(C | A) = 0.5 as expressing doxastic indifference, because if

$$\tau(C \mid A) = 0 \approx \log_a\left(\frac{P(Y=0 \mid X=1)}{P(Y=1 \mid X=1)}\right) = 0 \leftrightarrow \frac{P(Y=0 \mid X=1)}{P(Y=1 \mid X=1)} = 1 \leftrightarrow P(Y=0 \mid X=1) = P(Y=1 \mid X=1) = 0.5 .$$

This point appears to not have been fully realized, when Spohn (2012: 107) suggests that there is no equivalent in the probability calculus for (11a)-(11d).

[57] Reference: Spohn (2012: ch. 8).

[58] On conditional odds: $O_i = \frac{P(Y=1|X_1,\ldots,X_n)}{1-P(Y=1|X_1,\ldots,X_n)}$ for $i = 1,\ldots,n$.

[59] Reference: see also Politzer & Bonnefon (2006).

[60] Reference: Eid *et al.* (2010: section 16.6), Howell (1997: ch. 9).

[61] Acknowledgement: in discovering this I was helped by the responses to my inquiry at a forum for statisticians: stats.stackexchange.com/questions/66430

[62] Proof: $e^{b_1} = \frac{\frac{P(Y=1|X=1)}{P(Y=0|X=1)}}{\frac{P(Y=1|X=0)}{P(Y=0|X=0)}} = \frac{P(Y=1|X=1)}{\frac{P(X=1|Y=0)\cdot P(Y=0)}{P(X=1)}} \cdot \frac{P(Y=0|X=0)}{\frac{P(X=0|Y=1)\cdot P(Y=1)}{P(X=0)}} = \frac{P(Y=1,X=1)}{P(X=1,Y=0)} \cdot \frac{P(Y=0,X=0)}{P(X=0,Y=1)}$

[63] Proof: $e^{b_1^*} = \frac{\frac{P(X=1|Y=1)}{P(X=0|Y=1)}}{\frac{P(X=1|Y=0)}{P(X=0|Y=0)}} = \frac{P(X=1|Y=1)}{\frac{P(Y=1|X=0)\cdot P(X=0)}{P(Y=1)}} \cdot \frac{P(X=0|Y=0)}{\frac{P(Y=0|X=1)\cdot P(X=1)}{P(Y=0)}} = \frac{P(Y=1,X=1)}{P(Y=1,X=0)} \cdot \frac{P(Y=0,X=0)}{P(Y=0,X=1)}$

[64] Same point in probabilistic terms: in probabilistic terms, the degree of perceived sufficiency can be cashed out in terms of how far above 0.5 P(Y=1|X=1) is, and the talk about degrees of perceived necessity can be cashed out in terms of how far below 0.5 P(Y=1|X=0) is.

[65] Responding to a potential objection: however, as Karl Christoph Klauer (personal communication) has pointed out, more changes will have to be made, if we are to ensure that a probability measure comes out in the end, where the probabilities sum up to one. The most natural changes are that the upwards adjustment of P(Y=0, X=1) is followed by a downward adjustment of P(Y=1, X=1) in the presence of disablers and that the upward adjustment of P(X=0, Y=1) is followed by a downward adjustment of P(X=0, Y=0) in the case of alternative antecedents. This is an issue that we will return to in sections IV 2.2.2 and IV 3.1.

[66] Extension: we will return to the issue of what predictions can be derived for AC and DA in section IV 3.1.

[67] Reference: cf. Oaksford & Chater (2010b).

[68] Acknowledgement: this useful way of Evans and Over's position is due to Karl Christoph Klauer (personal communication).

[69] Reference: (personal communication).

[70] Reference and an important qualification: e.g. Oaksford & Chater (2003), Oberauer and Wilhelm (2003), Evans, Handley, & Over (2003), Overauer, Weidenfeld, & Fischer (2007). One important thing to note about this evidence, however, is that officially Acc(if A, C) = P(C|A) is part of the suppositional theory of conditionals and not the equation P(if A, C) = P(C|A) due to worries about conditionals not expressing propositions (cf. section II 1.2). Yet, as Douven (forthcoming: ch. 3) shows, it is actually the latter and not the former that has been substantiated by the data.

[71] An alternative: the meta-linguistic approach in chapter II makes it natural to expect P(if A, C) = P(A is a reason for C), so these two candidates could be contrasted empirically. However, it should be noted that the latter approach was an optional extension of the ranking theoretic approach to conditionals, which was introduced, because it allowed us to account for compositionality and the perceived objective purport of indicative conditionals.

[72] Explication: this is a further qualitative prediction that I attribute to Spohn (2013a) in addition to the one we saw in section 2.3.

[73] Explication of enabling conditions: the presence of oxygen is an enabling condition for striking a match being rank raising for the lightening of the match. The reason why enabling conditions for A being a reason for C have to be excluded is that otherwise it is no longer the case that A would raise the two-sided rank of C above 0, if B takes the value 'false'.

IV

The Logistic Regression Model and the Dual Source Approach[74]

Abstract: The main purpose of this chapter is to consider the implications for the dual source approach of the model of conditional reasoning put forward in the last chapter. A secondary aim is to explicate more generally the extent to which the predictions derived there either fit with existing empirical findings, or the extent to which they make up unique predictions of the ranking-theoretic approach to conditionals. Finally, in accordance with the methodological recommendations advanced in chapter I, it is to be investigated, whether any of these predictions count as hard to vary.

1. Introduction

Before jumping right to the comparative discussion, a short exposition of what is to come is given below.

We start out in section 2 with a general introduction to the dual source approach. In section 2.1, a mathematical implementation of it is considered. The purpose of section 2.2 is then to evaluate the prospect of extending the dual source approach with the logistic regression model from chapter III as a model of the content mode of reasoning. In section 2.2.1, a comparison is made between the latter and the Oaksford & Chater model used in Klauer *et al.* (2010). In section 2.2.2, we consider how the presence of the rule in the conditional inference task can be modeled by means of the logistic regression model. In section 2.3, the comparative discussion is extended to covering the newest development of the dual source approach; to wit, the INUS theory advanced by Klauer.

In section 3, we then finally take up the question of implementing the methodological recommendations from chapter I, when the prospects of identifying unique and hard to vary predictions among the predictions derived in chapter III is discussed. In section 4, we briefly take up one possible objection against the use of regression models in general to model judgment made by Gigerenzer and the ABC group. Finally, an appendix has been attached to deal with the problem of what is learned by conditional information in a way, which meets Douven's (2012) criteria of adequacy.

2. The Dual Source Approach

In Beller & Spada (2003) one finds an attempt to strike a balance between domain-general, deductive reasoning and domain-specific, content-based forms of reasoning by emphasizing that reasoning draws on two distinct sources of inferences: a content based source and a form based one.

This approach has been substantiated by a series of subsequent experiments dealing with deontic conditionals (e.g. Beller (2008)), deductive inferences, and probabilistic

inferences (e.g. Beller & Kuhnmünch (2007), Klauer, Beller, & Hütter (2010), and Singmann & Klauer (2011)).

A central methodological innovation in these experiments is the use of a *rule-free, baseline condition* in the conditional inference task that we encountered in chapter III. To refresh, the conditional inference task consists of evaluating the conclusion of the following inferences: MP (*modus ponens*: $p \rightarrow q, p \therefore q$), MT (*modus tollens*: $p \rightarrow q, \neg q \therefore \neg p$), AC (*affirmation of the consequent*: $p \rightarrow q, q \therefore p$), and DA (*denial of the antecedent*: $p \rightarrow q, \neg p \therefore \neg q$). The point of introducing the base-line condition is to measure pure content-based reasoning and experimentally isolate the contribution of the conditional rule in these inferences. This is done by comparing the participants' performance on tasks, where the conditional rule is present, with their performance on tasks, where they have to evaluate either the probability or necessity of the conclusion given the minor premise alone (e.g. evaluating 'q' on the basis of 'p' alone in *the reduced MP problem*). In this way, the proponents of the dual source approach hope to isolate the respective contributions of the content mode of reasoning and the form mode of reasoning. The underlying assumption is that performance in the baseline condition will tap into the content mode of reasoning, when realistic material (e.g. causal statements) is used, whereas the presence of the conditional rule will activate the form-based mode of reasoning.

This is but one of the methodologies that have been used to empirically validate the dual source approach. In general, it is supported by results showing an effect for variation in the logical form of the arguments in experimental conditions, where the content stays invariant, and an effect for variation in content in experimental conditions, where the logical form remains invariant.

One of the major contributions of the dual source approach consists in challenging Oaksford & Chater's (2007) attempt to completely account for reasoning in terms of Bayesian models that only work with a content-based component and attempt to reinterpret the performance on tasks used to assess deductive competence by means of probabilistic models. In Klauer *et al.* (2010) such an attempt was compared with the dual source approach on the conditional inference task and the results indicate that despite the documented poor logical performance, there seems to be some sensitivity to the syntactical

form of the arguments that interacts with the content mode of reasoning, which needs to be accounted for. This finding is in turn reinforced by Singmann & Klauer's (2011) finding that there is a double dissociation showing that giving deductive instructions[75] increase acceptance of valid but implausible problems and giving inductive instructions[76] increase acceptance of invalid but plausible problems.

Both of these findings thus indicate that it would be premature to work with models that only emphasize one source of inferences. Yet, in the latter results we at the same time see that this doesn't mean that the participants are better at solving logical problems than originally thought, insofar as the effect in Singmann & Klauer (2011) was only found in relation to affirmation problems (MP and AC) due to the subjects' difficulties with MT. Indeed under most conditions, the valid MT was not endorsed more than the logically invalid DA under deductive instructions (and in experiment 1 it was actually reverse).

Furthermore, in several studies *a reformulation task* has been introduced to test, whether the subjects would be able to integrate content information into a complete logical representation of the task (e.g. in Beller & Kuhnmünch, 2007). As a model of the subjects' *complete interpretation of a task* involving causal conditionals, Beller & Kuhnmünch (2007) suggest that alternative antecedents can be represented as disjuncts, and disabling conditions can be represented as conjuncts, in bi-conditionals that exhaustively outline all the possible causes. (Or rather: what the participant takes to be an exhaustive enumeration of all the possible causes.) So if we take the causal conditional 'If a car is involved in an accident, then the airbag of the car inflates' with an oversensitive sensor as an alternative antecedent, and an insensitive sensor as a disabling condition, it is to be represented in the following form:

$$((\text{accident}_{\text{Car}} \wedge \neg\, \text{insensitive\_sensor}_{\text{Car}}) \vee \text{oversensitive\_sensor}_{\text{Car}}) \leftrightarrow \text{airbag\_inflation}_{\text{Car}} \quad (1)$$

What the participants' performance on this task shows is that they do have some competence in picking out syntactically appropriate descriptions for the conceptual relations activated by their background knowledge.

As the dual source approach has been formulated, it is in principle capable of being combined with various theoretical frameworks. On the one hand, this lack of commitment is a strength as it enables many different applications. On the other, it simultaneously means that both the content component and the form component are in need of further specification. Throughout this chapter, we shall therefore examine the prospects of expanding the dual source approach by means of the logistic regression model. But first, we will take a look at a mathematical implementation of the dual source approach in the next section.

2.1 A Mathematical Implementation

In Klauer *et al.* (2010), a mathematical model is put forward to explicate the dual source approach, which generates quantitative predictions for conditional reasoning using probabilistic instructions. The model takes the following form:

$$\lambda\{\tau(x)+(1-\tau[x])\cdot\xi(C,x)\}+(1-\lambda)\cdot\xi(C,x) \qquad (2)$$

Where the 'τ' parameter quantifies on a probability scale the subjective certainty that the conclusion must be accepted/rejected on logical grounds for each of the four types of inferences (i.e. τ(MP), τ(MT), τ(AC), τ(DA)). In contrast, the 'λ' parameter quantifies the respective weight attributed to the form mode of reasoning and the content mode of reasoning, and the 'ξ' parameter quantifies on a probability scale the degree to which the conclusion in the inference in question is taken to be consistent with background knowledge for each particular content, C.[77] Hence, for each content, C, four ξ parameters are estimated (i.e. $\xi(C_1, MP)$, $\xi(C_1, MT)$, $\xi(C_1, AC)$, $\xi(C_1, DA)$).

One thing to notice about this model is that the final judgment is treated as the outcome of either perceiving the conclusion as correct by means of the content mode of reasoning (with the weight of 1-λ), or as correct by the form mode of reasoning (with the weight of λ). In the case the form mode of reasoning fails to reach a decision about the logical correctness of the conclusion, use of the content mode of reasoning is again treated as a fallback option with a probability of (1-τ).

To model the content mode of reasoning, the following model formulated by Oaksford & Chater is used as a substitute for $\xi(C_i, MP)$, $\xi(C_i, MT)$, $\xi(C_i, AC)$, $\xi(C_i, DA)$ to reduce the number of free parameters in equation (2) to three free parameters for each particular content, C:

$$(MP_R) \qquad P(q|p) = 1 - e(C) \tag{3}$$

$$(MT_R) \qquad P(\neg p|\neg q) = \frac{1-b(C)-a(C)e(C)}{1-b(C)} \tag{4}$$

$$(AC_R) \qquad P(p|q) = \frac{a(C)(1-e[C])}{b(C)} \tag{5}$$

$$(DA_R) \qquad P(\neg q|\neg p) = \frac{1-b(C)-a(C)e(C)}{1-a(C)} \tag{6}$$

Where the parameter a(C) is the perceived probability of p events for the content C, b(C) is the perceived probability of q events for the content C, and e(C) is the conditional probability of *not-q* given p for the content C.

Klauer *et al.* (2010) then conducted four experiments to compare the performance of (2) with an attempt of modeling the data by (3)-(6) alone. That is, the following two models were contrasted: (i) a dual source approach that takes the form of equation (2) and uses equations (3)-(6) to model the contribution of the content mode of reasoning, and (ii) an attempt to model the participant's performance on the conditional inference task using equations (3)-(6) alone with the idea being that both the reduced inference problems (without the conditional rule) and the standard conditional inference problems (with the conditional rule) access the participant's background knowledge in the form of conditional probabilities. In the latter case, the effect of the presence of the conditional rule is predicted to consist in a reduction to the value assigned to the exceptions parameter, e(C).

The conclusion they reached was that both models performed well. Yet, the first model had an advantage in terms of the fit of the model to the data relative to the number of free parameters employed, and the capacity of the model to map its estimated, free parameters onto the experimental manipulations in a way that made sense qualitatively.

For the purposes of introducing the logistic regression model as a possible replacement of equations (3)-(6) in the dual source approach, it is important to distinguish between two possible contributions of the form mode of reasoning. In principle it is either possible that the form mode of reasoning enters the picture merely to process the conditional rule (contribution$_1$), or that the form mode of reasoning is involved in analyzing, whether the combination of premises and conclusion instantiates a valid argument schema (contribution$_2$). (And, of course, the two can also be combined and interact with the content mode of reasoning.)

The reason why we need this distinction is that as the logistic regression model is able to account for the effect of adding the conditional rule (as we shall see in section 2.2.2), it might be argued that a form based mode of reasoning that goes beyond what can be modeled by it only enters the picture in relation to the computations involved in contribution$_2$.

This opens up for the question of whether there would be a difference in the contribution of form mode of reasoning depending on variations of the conditional rule. That is, if the participants were presented with changes to the conditional premise in MP, MT, AC, and DA, would they then have enough form competence to make the appropriate changes? Klauer *et al.* (2010) actually implemented such a manipulation in experiment four, when they used the so-called negations paradigm and introduced the following conditional rules: 'if *p* then *q*', 'if *p* then ¬*q*', 'if ¬*p* then *q*', and 'if ¬*p* then ¬*q*'. Moreover, in another experiment they contrasted 'if *p* then *q*' with 'only *p* if *q*'. Hence, the findings that these manipulations induced may be counted as documenting the psychological reality of contribution$_2$.

But which semantics of conditionals we adopt will, of course, end up having repercussions for how we conceive of contribution$_2$. In particular, present attempts to fit equation (2) to the data have all ended up estimating values for τ(AC) and τ(DA) that are greater than zero. Yet, if either the suppositional theory of conditionals or the horseshoe analysis of the mental model theory is adopted, then the subjects are thereby depicted as utilizing a form based mode of reasoning that makes them perceive invalid argument schemes as valid (!)

However, we should probably refrain from such a conclusion as we have already seen in section III 3.3 how both AC and DA have legitimate uses not captured by these theories. In contrast, on a relevance approach all MP, MT, AC, and DA are to be treated as valid, although they are to be endorsed to different degrees. So if $contribution_2$ encompassed the analysis of reason relations, we would not be forced to attribute irrationality to the participants on this score (which would surely be a welcome result). What this would suggest is that there may be a layer to the form based mode of reasoning that goes beyond Boolean algebra by requiring the comprehension of reason relations. Formulated in terms of the terminology introduced in chapter II, decoding the dialectical compositionality of arguments through the expressed reason relations would have to be included among the tasks performed by our form mode of reasoning. If the ranking-theoretic approach to conditionals is on the right track, then we already see what this amounts to in relation to the connective 'if then'. But the discussion in chapter II pointed to many other expressions that clearly serve an argumentative function as well.

2.2 Using the Logistic Regression Model to model the Content Component

Turning to the content-based component, the hypothesis that we shall be entertaining in this chapter is that we can use the logistic regression model to give us an account by substituting it for Oaksford & Chater's model.

According to this suggestion, one paradigmatic contribution of the content mode of reasoning consists in the understanding, use, and decoding of predictor-relationships communicated (as emphasized by conditionals containing the predictive modal 'will'). That is, when using paradigmatic examples of indicative conditionals, the speaker is expressing a predictor relationship between the antecedent and the consequent, and the expectation that the hearer forms, when hearing such conditionals, is that there is a predictor relationship between the two. Accordingly, what the hearer does, when acquiring new information through others' use of conditionals is to set the regression weights of the assumed predictor relationship between the antecedent and the consequent to default values, which

are subject to modification through linguistic qualifiers (e.g. 'if Obama wins the election, then it is *highly likely* that apartheid won't be reintroduced in North America').

Furthermore, when engaging in conditional reasoning, the subjects will be thought of as exploiting their understanding of the predictor relationship between the antecedent and the consequent to draw inferences about how likely it is that one relatum of the predictor relationship is true/false given that the other is assumed to be true/false.

Both of these functions thus emphasize the tight connection between the use of conditionals and the expression of predictor relationships. However, the reference to the dialectical compositionality of arguments above reminds us of the central role that conditionals have of making reason relations explicit in argumentative contexts, so that they can be made the target of justificatory challenges (cf. chapter II). The reason why the present account is capable of encompassing both of these roles is that on the notion of epistemic relevance utilized in the ranking-theoretic account of conditionals, predictor relationships and reason relations end up being two sides of the same coin.[78]

To explore these various ideas further, section 2.2.1 opens up for a comparative discussion with the Oaksford & Chater model and section 2.2.2 discusses the prospects of modeling the presence of the conditional rule in the conditional inference task using the logistic regression model. And finally, appendix 3 expands on the suggestions above about what is involved in learning conditional information by using the results from section 2.2.2 to show how we can meet Douven's (2012) criteria of adequacy.

2.2.1 Comparing the Logistic Regression Model and Oaksford & Chater's Model

As the present suggestion is in effect to replace the Oaksford and Chater model in the dual source theory with the logistic regression model, it is only fitting that we compare the two. More specifically, a potential objection made by Klauer and Singmann (personal communication) needs to be encountered: using the logistic regression model to model the data from the baseline condition in experiment one in Klauer *et al.* (2010), they noticed that the two models were both reparametrizations of the joint probability distribution, and that

in fact they had an identical determination coefficient of 0.97 (which means that both models account for 97% of the variance in the data).

Due to this observation, it may appear that it doesn't matter much which model we pick as all the predictions of the logistic regression model could be derived from the Oaksford & Chater model and *vice versa*. Hence, if the determination coefficient in the rule-free baseline condition and the number of free parameters were the only criteria of adequacy that we had to rely on in comparing these two models of the content mode of reasoning, it would be impossible to discriminate between the two.

But when viewed from a more general philosophy of science perspective, it can be seen that these are *not* the only criteria of adequacy with which we can assess the qualities of a theory. Further candidates include: (1) the psychological plausibility of the parameters of the model, (2) the capacity of the model to generate new, interesting research questions, (3) the capacity of the model to throw new light on existing theoretical puzzles, and (4) the ability of the model to inspire researchers to derive new, interesting predictions. When we compare the two models along (1)-(4), they are far from equal, as I will now go on to argue.

In considering the question of the psychological plausibility of the parameters used in the two models by considering the values of the estimated parameters as information that summarizes the part of the reasoner's knowledge about the particular content that is used in conditional reasoning, the main difference is the inclusion of relevance considerations in the present model. This difference is a symptom of the fact that Oaksford & Chater's model is built on the suppositional theory of conditionals, whereas the logistic regression model is built on Spohn's (2013a, forthcoming) theory of conditionals. As a result, the latter allows us to model the presence of the rule in the conditional inference task in terms of an increase in the perceived relevance of the antecedent for the consequent (cf. section 2.2.2) in accordance with the relevance approach to conditionals motivated in chapter II. In contrast, the Oaksford & Chater model would model the presence of the rule through changes to the exceptions parameter e(C).[79] So here we see an instance of how a difference in the psychological interpretation of the model's parameters has implications for further applications of the models (in spite of the mathematical equivalence). Moreover, as a result of this difference, the present model allows us to establish continuity to Rescorla's (1988)

work on classical conditioning, where relevance considerations likewise enter the picture under the heading of 'contingency', as we have seen in chapter III.

In relation to comparing the two models on their capacity to raise new, interesting research questions, the reader is deferred to the points just made in section 2.2 as well as to the comparative discussion of the suppositional theory and the logistic regression model in section III 3.6, where it was emphasized that it is hard to see the evolutionary point of using the conditional as a linguistic device to simulate possibilities unless the possibilities simulated are constrained by relevance considerations. Moreover, the latter section also shows that the logistic regression model is capable of throwing new light on existing, theoretical puzzles through its suggestion about the computational task involved in performing the Ramsey test. Furthermore, the capacity of the model to rationalize the endorsement rates of MP, MT, AC, and DA that don't fit the normative model based on deductive logic speaks in favor of it. As we have seen, although neither deductive logic nor the suppositional theory validate AC and DA, they have very sensible applications in argumentative contexts and in abductive reasoning (cf. section III 2.3).

Finally, by including a parameter for representing perceived relevance, the model allows us to use Spohn's taxonomy of reason relations to derive a range of nice predictions (some of which match existing psychological findings and some of which are novel). At this point it might be objected that one should in principle be able to derive the same predictions from the two models as they are both reparametrizations of the joint probability distribution. Yet, although this may be formally correct, these predictions would not be theoretically motivated by the semantics of conditionals that Oaksford & Chater's model is based on. Moreover, we may also notice that historically it seems that no one has actually used Oaksford & Chater's model to derive the predictions in question. This is probably no accident. For although it should be *formally possible*, we also need to take into account the fact that the researchers, who are supposed to derive the predictions in question, are guided by the research questions that the parameters of the model inspire, and that it might not be as easy to discover these predictions on the basis of either model. This observation about the psychology of the researchers is not only interesting in itself. But it also acquires importance for philosophy of science more generally, because it implies

that two formally equivalent models need not be equal in their contribution to the theoretical discourse, if they inspire the researchers to investigate different predictions and raise different research questions.

I take it that all of these points count in favor of the logistic regression model over against Oaksford & Chater's model.

2.2.2 Comparing the Logistic Regression Model and the Dual Source Model

The purpose of this section is to consider some of the ways in which the logistic regression model could itself be used directly to model the effect of the presence of the rule found in Klauer *et al.* (2010). That is, we will now consider the possibility of using the logistic regression equations as not only providing a model of the content mode of reasoning in equation (2) but as a direct competitor of equation (2).

Before we begin, we must notice that since we already have the results, methodologically we are in a peculiar situation, where we are designing a model to retrodict known findings rather than predicting them in advance. It is always easier to modify a model to existing findings. Hence, such "predictions" should not be allowed to count much in favor of the model, *unless* they are based on *hard to vary* aspects of the underlying theory (as opposed to be based on auxiliary hypotheses).

On the face of it, the main finding in Klauer *et al.* (2010) was that the endorsement rates of MP and MT were boosted in the presence of the rule. The logistic regression model would be able to account for these findings, if we assume that the effect of introducing the rule is not merely to raise b_1 (i.e. to increase expectations about the epistemic relevance), but specifically to increase the perceived sufficiency (i.e. increasing b_1 and decreasing b_0^*). For in section III 3.5, the latter has been shown to have the effect of increasing MP and MT. However, whereas the former effect is straightforwardly motivated by the ranking-theoretic approach to conditionals, it might seem that the latter lacks a principled justification.

We will now show that this is not the case. In Spohn (forthcoming) it is said that 'the circumstances are such that' reading of conditionals is the most interesting, expressive

function of the 8 main expressive functions he analyzes. As it turns out, this reading also lays the foundation for his account of causal conditionals, which happens to be the kind of stimulus material used in Klauer *et al.* (2010). The idea behind Spohn's (2013a, forthcoming) analysis of 'the circumstances are such that' reading of conditionals is that the positive, epistemic relevance of the antecedent for the consequent is itself based on a condition, which is assumed to obtain, when the conditional is asserted. As a result, the speaker is portrayed as expressing an *unconditional belief* about the fulfillment of this condition (in addition to the features of his *conditional beliefs* expressed by the conditional itself) about which there can be a factual dispute. In the case of causal conditionals, this takes the form of having an unconditional belief about the actual history being such that the antecedent was somehow required to bring about the consequent. Epistemically, what this means is that the positive relevance of the antecedent for the consequent is itself grounded in the actual history up until the occurrence of the consequent at time t^1 (where we have excluded the occurrence of the antecedent at time t). So according to this analysis, causes can be thought of as "reasons given the history" (Spohn, forthcoming).

Paraphrased in terms of the logistic regression model, we can understand 'history not being as it actually was up until the effect occurred' as a *disabler* for the positive relevance of the antecedent for the consequent in causal conditionals. Accordingly, 'history being as it actually was up until the effect occurred' can be analyzed as *the non-occurrence of a disabler*, which we might demarcate terminologically as *an enabling condition*. Hence, the unconditional belief that causal conditionals express, according to Spohn's analysis, is that the enabling condition for the epistemic relevance of the antecedent for the consequent is fulfilled. And as we have already seen in chapter III how the absence of disablers has the effect of increasing MP and MT, it turns out that we are able to account for the main finding of Klauer *et al.* (2010) after all, once we apply Spohn's account of causal conditionals to their stimulus material (which consisted of … causal conditionals). That is, the conditional rule raises the perceived sufficiency of the antecedent for the consequent in virtue of conveying the information that an enabling condition for the positive, epistemic relevance of the antecedent for the consequent is fulfilled. Hence, it turned out to be possible to give a principled account of the main effect in that study after all.

However, at this point, we need to return to a possible objection (cf. endnote 65). In chapter III, it was said that the effect of the *absence* of disablers is to lower P(X=1, Y=0) and the effect of the *absence* of alternative antecedents is to lower P(X=0, Y=1). However, more changes will have to be made to the probability distribution to ensure that we end up with a probability measure in the end, where the probabilities sum up to one.[80] The most natural changes are that the downward adjustment of P(X=1, Y=0) is accompanied by an upward adjustment of P(X=1, Y=1) in the *absence* of disablers and that the downward adjustment of P(X=0, Y=1) is accompanied by an upward adjustment of P(X=0, Y=0) in the *absence* of alternative antecedents.

This extension does not affect the predictions derived in chapter III. But it has implications for what further predictions can be derived on the basis of the logistic regression model. In particular, it implies that the increases to the perceived sufficiency affected by the absence of the disablers will also raise AC in addition to MP and MT and that the increase to MP should be larger than the increase to MT as: (a) P(X=1, Y=1) figures in the equation for AC, and (b) MP will be affected by both the changes to P(X=1, Y=1) and P(X=1, Y=0).

To make this more perspicuous, the equations for the parameters from chapter III have been inserted in the logistic regression model in the equations below (and terms that cancel out have been removed):

$$MP_R: \frac{1}{1+e^{-(ln[P(X=1,Y=1)]-ln[P(X=1,Y=0)])}} \quad (7)$$

$$MT_R: \frac{1}{1+e^{ln[P(X=1,Y=0)]-ln[P(X=0,Y=0)]}} \quad (8)$$

$$AC_R: \frac{1}{1+e^{-(ln[P(X=1,Y=1)]-ln[P(X=0,Y=1,)])}} \quad (9)$$

$$DA_R: \frac{1}{1+e^{ln[P(X=0,Y=1)]-ln[P(X=0,Y=0)]}} \quad (10)$$

Keeping this prediction in mind, we now inspect the results cited in Klauer *et al.* (2010) for a second time. For the results reported in their figure 1 and 3, it holds that MP is

boosted more than MT and that AC is boosted in the presence of the rule. This pattern fits with our predictions. For the results reported as figure 4 and 5 in Klauer *et al.* (2010), we see essentially the same pattern with the only difference being that there are also now some minor increases to DA as well. As these changes to DA were not consistent across all experiments, it appears in comparison to be a more negligible effect, which should be accounted for in terms of auxiliary hypotheses.

The following candidates suggest themselves: (i) that the participants sometimes fail to take alternative antecedents into account, because they are engaging in what has been called *closed-world* reasoning (cf. Beller & Kuhnmünch, 2007), where they assume that they have taken all the relevant factors into account, when they have processed both the major and the minor premise in evaluating the conclusion of MP, MT, AC, and DA. Or (ii) that the participants sometimes fail to take alternative antecedents into account due to cognitive limitations. Generating alternative antecedents, and taking their contribution to the problem at hand into account, is itself a process that takes up cognitive resources. So perhaps the participants are already so preoccupied with processing the major and minor premises in the conditional inference task that they simply fail to invest further cognitive resources in initiating the former process (although they could easily be brought to realize that they should have known better). The second auxiliary hypothesis is plausible, because it is known from a number of experiments that ordinary subjects tend to be 'cognitive misers' that will substitute resourceful analytical thinking for rough heuristic approximation whenever they can (Stanovich, 2011: 21, 29, Kahneman, 2012: ch. 5). But clearly, the two auxiliary hypotheses can also be combined.

However, in accordance with our introductory remarks to this section, such explanations that take a recourse to auxiliary hypotheses should not be taken to count much in favor of the ranking-theoretic approach to conditionals (what they do, if successful, is rather to show that the data need not be considered as evidence against the latter account). Yet, what does count in favor of our model is its ability to retrodict the pattern that MP is boosted more than MT and that AC is boosted in the presence of the rule, insofar as it has received a principled justification.

2.3 The INUS Theory

The INUS theory is an attempt by Klauer (manuscript) to use John Mackie's analysis of causes as prior INUS conditions (i.e. insufficient, but non-redundant parts of unnecessary but sufficient conditions for the effects) as the basis of a specification of the content mode of reasoning in the dual source approach. The theory builds on Beller & Kuhnmünch's (2007) approach to causation, but goes beyond it at some crucial junctures, and we shall therefore briefly raise some comparative points about it in the following.

Klauer (manuscript: 2) gives the following example of an INUS condition:

> A spark causing a fire is not a sufficient cause for the fire; it causes a fire only in conjunction with other conditions such as the presence of easily inflammable materials X_1 nearby, the absence of water X_2, and so forth.

I take the point to be that the antecedent event (i.e. a spark occurring) is by itself an *insufficient condition* for the consequent event (i.e. the fire occurring), because there are other conditions like the presence of easily inflammable materials (X_1), which are (what we have been calling) it's enabling conditions. However, the antecedent event is a *non-redundant* part of a sufficient condition as the enabling conditions by themselves are incapable of bringing about the consequent event. Yet, the antecedent event is still only part of an *unnecessary, sufficient condition*, when all the enabling conditions are fulfilled to the extent that there can be alternative antecedents, which are likewise capable of bringing the consequent event about (when their respective enabling conditions are likewise fulfilled). As this explication makes clear, there is thus a close connection between the INUS theory and Spohn's (2013a) explication of "the circumstances are such that" reading of conditionals in that both formally represent the enabling conditions as conjuncts or intersections in the antecedent clause.

The theory adopts the same formal representation of the total cause of an event as the one we have already encountered from Beller & Kuhnmünch (2007):

$$((\text{accident}_{Car} \wedge \neg\, \text{insensitive\_sensor}_{Car}) \vee \text{oversensitive\_sensor}_{Car}) \leftrightarrow \text{airbag\_inflation}_{Car} \quad (1)$$

But it introduces the innovation with respect to the latter that the list-like enumeration of disablers and alternative antecedents is allowed to be gappy with dots filling out the empty spaces. Klauer (manuscript) moreover suggests another measure to model the subjects' incomplete knowledge: that the disablers and alternative antecedents are typically left implicit. Hence, even the gappy list is thought as the result of explication, which only occurs under special circumstances.

However, since the theory nevertheless follows Beller & Kuhnmünch (2007) in adopting the logical equivalence in (1) as an analysis of how the information required to perform the conditional inference task is organized, a comparison with the latter and the logistic regression model will also carry over to a comparison with the INUS theory as advanced by Klauer.

In particular, we shall follow Markman's (1999) "shoppers' guide to knowledge representations for psychologists" in comparing the different representational formats on their expressive power. Of special interest in this regard are the following observations:

1) The logistic regression model uses its regression weights to quantify the contribution of less than perfect predictors, whereas equation (1) can only represent cases of deterministic causation, where it holds that, whenever one of the antecedents (and its enabling conditions) obtain so does the consequent.

2) The output of the logistic regression model is a conditional probability, whereas the output of equation (1) is a truth value.

3) In contrast to equation (1), the logistic regression model is capable of representing the effect of disablers as consisting in merely lowering the regression weight of the predictor, without making their effect an all or nothing affair. (In principle, one could also use the logistic regression model to allow for disablers to vary in strength depending on the context.)

4) In contrast to equation (1), the logistic regression model allows us to represent perceived sufficiency and necessity as something that comes in degree. (Cf. section III 3.4)

To illustrate the importance of these comparative points: in no experiments that I am aware of do the endorsement rates of either MP & MT or AC & DA decrease to zero, when the subjects are supplied with information about disablers or alternative antecedents. Yet, in contrast to the logistic regression model, equation (1) lacks the expressive power to represent such gradual changes to the perceived sufficiency and necessity. Of course, this point and the other observations outlined in 1) - 4) are just symptoms of the fact that (1) is a *logical representation* of the information, which only represents *full beliefs*, whereas the logistic regression model provides us with a *probabilistic representation*, which allows us to represent *degrees of belief*. It would thus be interesting if a probabilistic version of the INUS theory is formulated.

A further central theme in comparing the INUS theory to the logistic regression model is the difference between *mere predictions* and *explanatory-based predictions*. The reason is that in the logistic regression model, we have a direct way of representing the effect of disablers and alternative antecedents through the way they modify the parameters of the model, without the subject having to enumerate the former in a list. It is thus possible to maintain on the basis of the logistic regression model that memory traces of alternative antecedents and disablers influence the computations by modifying the corresponding parameters, without the participants having to produce a declarative representation of the corresponding alternative antecedents and disablers.

To take an example, not knowing much about car engines, it is possible for Joe to adjust the weight that he would assign to „turning the key" as a predictor of the car starting due to an experience of its occasional failure, without having to worry about why the engine failed under those circumstances.

So although the disablers may also figure as *explanatory reasons* of why the engine failed to start, it is possible for Joe *not* to produce the corresponding explanations and still be able to accommodate their influence in the cognitive processes involved in prediction. As far as prediction goes, the subject need not in the first instance know that Z_1, Z_2, and Z_3 are disablers for using X_0 to predict $Y = 1$. It suffices that he adjusts his expectation for the ability of $X_0 = 1$ to predict $Y = 1$, whenever he encounters obstacles like Z_1, Z_2, and Z_3.

In contrast, on the INUS model, the subject is already depicted as engaging in causal explanation, when using conditionals and conditional reasoning incorporating disablers and alternative antecedents. However, while it may be true that we do this as well, it is not obvious that there is not a level of prediction, where disablers and alternative antecedents are taken into account, without the participants being able to (or needing to, if cognitive resources are an issue) construct the corresponding explanations. Indeed, it might be possible to dissociate these two competences by giving participants tasks under cognitive load, or by introducing novel (artificial) stimulus domains, where they don't yet have any opinions about causal mechanisms. Furthermore, it might be that different experimental tasks tap into these two competences. If, for instance, one asks the subjects to justify, why they endorse MP to a lesser degree in regard to some (low-sufficiency) conditionals in comparison to other (high-sufficiency) conditionals, the subjects might be engaging in the resourceful task of enumerating possible disablers in the former case. If, on the other hand, they have to make a decision under cognitive load, or time pressure, it might suffice to take into account that there are some disablers (by adjusting the regression weights), without being able to make them explicit in even a gappy list.

As this suggests, it is possible that one would ultimately have to combine the logistic regression model with a gappy list representational format for modeling cases, where the participants have made *a first intuitive judgment* about the conclusions in the conditional inference task, and the subject is being called upon to *justify it afterwards* by retrieving examples of disablers and alternative antecedents. Such a synthesis would be also be attractive for the further reason that the INUS theory advanced in Klauer (manuscript) is a rich framework that makes a host of other predictions not commented on here.

However, to motivate that there is indeed such a level of sheer prediction, it is instructive to connect the present considerations with those that Fernbach & Erb (2013: 1329) make in relation to what they call 'the conditional probability theory of conditional reasoning', which they, *inter alia*, attribute to Oaksford & Chater:

> it might be possible in some cases to estimate conditional probability by calculating the conditional frequency of events in memory without thinking of specific disablers (e.g., counting memories of a gun failing to fire and dividing by the total number of attempts to fire a gun).

They here make the point that disablers can have an influence on the computations by affecting estimates of conditional frequencies of events in memory. This might be a useful way of thinking about the process of sheer prediction. What the logistic regression model adds to this is then that what the subjects are thereby doing is attempting to come up with an estimate of the model's parameters. More specifically, in order to estimate b_0, they would need to consider the frequency of 'Y=1' when 'X=0' is the case, and to come up with an estimate of b_1, they would need to consider the difference between the frequency of 'Y=1' when 'X=1' and 'X=0' is the case. This in turn gives us a grip on something that they hold that Oaksford and Chater's theory is silent on; to wit: "where judgments of conditional probability come from" (p. 1330).

But notice that epistemic relevance is defined in terms of conditional ranks:

$$b_1 = \tau(C \mid A) - \tau(C \mid \overline{A}) \approx ln\left(\frac{\frac{P(Y=1|X=1)}{P(Y=0|X=1)}}{\frac{P(Y=1|X=0)}{P(Y=0|X=0)}}\right)$$

Hence, this suggestion would be more or less worthless, unless the estimate of epistemic relevance is in itself the outcome of a more basic cognitive process. This is why the suggestion was added in chapter III that a fast and frugal heuristic should ultimately be developed to provide an approximation of the logged odds ratio. Laura Martignon (personal communication) has suggested that a heuristic that provides an approximation based on natural frequencies might be a feasible option and I tend to agree. However, these are all topics that await further research.

3. Unique and Hard to Vary Predictions of the Logistic Regression Model

The methodological recommendation in chapter I on how to make philosophical theories useful for scientific purposes was that philosophers could help bridging the gap by contributing to deriving predictions from their own theories. Its main motivation was that the philosophers should contribute to reaching the scientific goal of excluding possibilities from our hypothesis space, instead of just populating it with further ideas that we don't know in principle how to get rid of again. The constraint of identifying unique and hard to vary predictions was then introduced as a gold standard for how to contribute to this end.

As chapter III has already shown how to derive predictions from the ranking-theoretic approach to conditionals, we have already made much progress with respect to implementing these recommendations. So at this point it is only appropriate that we take it one step further and investigate, whether it is also possible to derive unique and hard to vary predictions from our logistic regression model.

As was already foreshadowed in chapter III, the potential for deriving *unique* predictions stems from manipulations along the relevance dimension and use of Spohn's taxonomy of reason relations. In relation to the latter, it was said that the categories of insufficient and supererogatory reasons are not conceptually distinguished from sufficient and necessary reasons in the empirical literature.[81] However, although they are not *conceptually distinguished*, it is still possible that experimental manipulations have been introduced, which implicitly targeted them, although the researchers in question were unaware of this fact.

In this context, it is useful to introduce the distinction that Politzer and Bonnefon (2006) have made between an *overt experimental paradigm* and a *covert experimental paradigm* in the conditional inference task. An example of the *overt paradigm* is Byrne's (1989) seminal paper, where an overt cue is introduced by the experimenter to explicit direct the attention of the participants to the existence of disablers or alternative antecedents, when making assessments of the endorsement rates of MP, MT, AC, and DA. In Byrne's example, the subjects were presented with 'If the library stays open then she will study late in the library'

as an additional (conditional) premise to the standard major and minor premises in modus ponens (i.e. 'If she has an essays to write then she will study late in the library' and 'She has an essay to write').

In contrast, in the *covert paradigm*, it is shown that disablers and alternative antecedents have an effect on conditional reasoning even when the participant is not explicitly reminded of them. An example is Cummins's (1995) seminal paper, where conditionals were pretested with a separate experimental group to fall in categories of many/few disablers and alternative antecedents. To illustrate, the conditional 'if Mary jumped into the swimming pool then she got wet' can then be used as case with many alternative antecedents, and the conditional 'if Joe cut his finger then it bled' can be used as a case with few alternative antecedents.

The contribution of Spohn's taxonomy of reason relations in this context is to sharpen our focus on, whether the disablers or alternative antecedents are also assumed to obtain by the participants as the logistic regression model produces different predictions for the cases, where: (1) the antecedent is an insufficient reason for the consequent (e.g. due to the obtainance of a disabler), (2) the antecedent is a sufficient reason for the consequent (which could be a case, where disablers are known to exist, but are assumed not to obtain in the present context), (3) the antecedent is a supererogatory reason for the consequent (e.g. due to the obtainance of alternative antecedents that are themselves sufficient reasons), and (4) the antecedent is a necessary reason for the consequent (this could also be a case, where alternative antecedents are known to exist, but they are assumed not to obtain in the present context). Hence, a reasonable goal for future experimental studies is to introduce the corresponding manipulations to both the overt and the covert paradigm to test the matching predictions of the logistic regression model.

When we turn to the potential for *unique* predictions stemming from manipulations of the relevance dimension, pretty much every prediction dealing with negative relevance and irrelevance will count as unique to the present theory as the relevance dimension is not normally taken into account in empirical work in experimental psychology. However, in experimentally investigating these predictions, it should be noted that care must be taken not to introduce general effects of negations as a conflating factor. The reason is that the

easiest way to experimentally introduce a negative relevance manipulation is to take a conditional with positive relevance and negate the consequent. But as other existing theories of conditionals already make predictions for negations (e.g. Oaksford & Chater, 2007), one would thereby face the problem of having to distinguish the effect of negative relevance from a general effect of negated consequents.[82] Hence, if the present theory is to count as having unique predictions relating to the negative relevance manipulation, it must be introduced experimentally in other ways than by negating the consequent of a positive relevance conditional (e.g. by picking the right content).

3.1 Identification of Hard to Vary Predictions

Having already identified various candidates for *unique* predictions of the logistic regression model, we now turn the question of whether any of its predictions count as *hard to vary*. The general motivation for requiring that some of its predictions be hard to vary is, as we remember, a concern with not only having experimentally distinguishable predictions of the novel theories we introduce to our hypothesis space, but also having predictions that would allow us to exclude them again. What we thereby mean to ban are attempts of safeguarding them with any number of *auxiliary hypotheses* that are not systematically related to the core claims of the theory.[83] In our context, the problem of identifying hard to vary predictions is exacerbated to the extent that the problem arises not only in relation to the logistic regression model, but also in relation to the ranking-theoretic approach to conditionals.

That is, even if we succeed in identifying hard to vary predictions of the logistic regression model, it might still be possible for proponents of the ranking-theoretic approach to conditionals to escape embarrassing challenges based on discordant empirical findings by dissociating themselves from the logistic regression model at strategically favorable times.

To start with the former issue, the goal of section III 3.5 was exactly to derive a set of predictions that holds for most values of the logarithm base of our regression equations. To this extent, these predictions certainly count as candidates for *hard to vary* predictions of the logistic regression model. Moreover, the existence of a few limiting cases, where they don't hold, need not concern us as section III 3.3 already introduced the constraint that the

choice of the logarithm base should be based on *independent support* in the form of evidence that targets the linguistic phenomenology of expressing degrees of beliefs. Whether these predictions also count as *hard to vary* predictions of the ranking-theoretic approach to conditionals will depend on how successful the case in chapter III was for translating negative ranking functions into logged probabilities with a logarithmic base, $a \in (0,1)$, where $a \neq \varepsilon$. As was noted in section III 2.1, this translation faces problems, when it comes to translating the sum of probabilities into the minimum of ranks.

As a result, we might only be justified in treating the logged probability translation as an *approximation* for when $a \neq \varepsilon$. Hence, the question arises, whether any predictions that can be derived using this approximation also count as *hard to vary* predictions of the ranking-theoretic approach to conditionals.

At this point, it is useful to observe that there is a methodological difference, when it comes to the use of formal models in mathematics (and mathematical philosophy) and the use of formal models in the empirical sciences. Whereas the former is driven by exactness and elegant derivations, the latter are more concerned with formulating useful models that behave nicely under most conditions. So for the latter, an approximation like ours may be legitimate for all practical intent.

But one issue is legitimacy. Another is whether we have found a way of preventing the proponents of the ranking-theoretic approach to conditionals to dodge unpleasant challenges by claiming in the face of incompatible evidence that any discrepancy between the empirical data and the model is due to the latter only serving as an approximation of ranking theory. In fact, I think that this problem can be dismissed, because the proponent of the ranking-theoretic approach to conditionals thereby incurs the obligation to show that any such incompatibility can then really be traced back to the translation's problems with translating the sum of probabilities into the minimum of ranks. If such attempts are unsuccessful, then any attempt of rescuing the theory on this behalf should probably be seen as *ad hoc* attempts of breathing air into a mistaken theory.

A perhaps more detrimental observation for the project of deriving hard to vary predictions from the ranking-theoretic approach to conditionals goes as follows. In chapter III the logistic regression model was put forward as a model of *the rule-free, baseline* condition

in the conditional inference task. Yet, it was supposed to help us derive predictions from Spohn (2013a, forthcoming), which is a theory of *conditional rules*. To be sure, the situation was amended in section IV 2.2.2, when suggestions were made for how to model the presence of the rule in a way that was systematically justified by the core claims in Spohn (2013a, forthcoming). However, there is still a legitimate worry that these measures of applying the ranking-theoretic approach to conditionals to the conditional inference task do not really form the basis of hard to vary predictions of the former. The underlying problem is that Spohn (2013a, forthcoming) employed a methodology of systematically investigating possible expressive functions of the conditional connective while remaining uncommitted about linguistic phenomenology and psychological applications. In the course of chapter III and IV, we have been struggling to explore a way of filling this gap in relation to the conditional inference task. But as these extensions have not been adopted as part of the official ranking-theoretic account, it seems that they can hardly count as hard to vary predictions of the latter. To deal with this problem, I see no other way than to encourage the proponents of the ranking-theoretic account of conditionals to adopt more determinate commitments with respect to these matters. Surely, the way the theory has been developed in chapter III and IV is not the only way of meeting this desideratum, but it at least constitutes one candidate.

In particular, in modeling the reduced conditional inference problems and the effect of adding the conditional rule, the present approach implicitly followed Brandom (1994, 2010: 44-8, 104) in holding that conditionals make reason relations explicit that can also be manifested in dispositions to draw content-based inferences without conditional rules. Accordingly, the reduced conditional inference problems was thought of as tapping into the participant's *perceived reason relations* as well and the presence of the conditional rule was depicted as merely serving to boost these. However, this way of extending the ranking-theoretic approach to conditionals may be optional, if other ways of modeling the expanded conditional inference task employed in Klauer *et al.* (2010) can be found.

3.2 Possible Exceptions

As the point of introducing hard to vary predictions is to make it possible to exclude the theory again from our hypothesis space, it is only appropriate, if we end this section with discussing two possible exceptions to the predictions of the logistic regression model that have been documented in the empirical literature.

More specifically, we saw with the correction introduced in section 2.2.2 that the *presence* of disablers should lead to an upward adjustment of P(X=1, Y=0) and a downward adjustment to P(X=1, Y=1), whereas the *presence* of alternative antecedents should lead to an upward adjustment to P(X=0, Y=1) and a downward adjustment to P(X=0, Y=0). On the basis of equations (7)-(10), the following predictions can be derived for the rule-free baseline condition of the conditional inference task based on these adjustments to the probability distribution:

> Presence of disablers: MP↓↓, MT↓, AC↓
>
> Presence of alternative antecedents: AC↓, DA↓↓, MT↓

Yet, Politzer and Bonnefon (2006: 486) suggest that both the *overt experimental paradigm* and the *covert experimental paradigm* encountered in the last section have produced the following pattern of results:[84]

> Thus, the two experimental paradigms concur to what we call here the Core Pattern of results: Disabling conditions defeat the conclusions of MP and MT (but usually not the conclusions of DA and AC) and alternative conditions defeat the conclusion of DA and AC (but usually not the conclusions of the valid MP and MT). This Core Pattern is endorsed by most if not all researchers in the field, and, apart from the occasional breach (see in particular Markovits and Potvin, 2001), has never been seriously questioned.

To be sure, the logistic regression model does succeed in capturing the result that disablers reduce the endorsement rates of MP and MT and that alternative antecedents reduce the endorsement rates of AC and DA, which is the main finding in the field. The

problem is the vague qualification in the parentheses that disablers *usually* don't influence AC and that alternative antecedents *usually* don't influence MT.

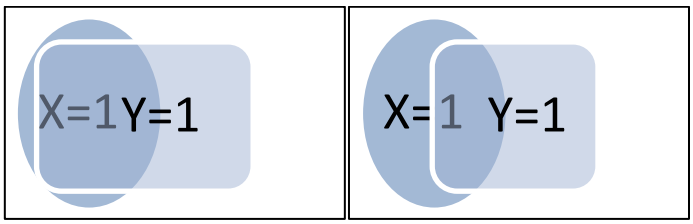
However, this problem is not one that the logistic regression model is alone in having. If the Oaksford and Chater model (i.e. equations (3)-(6)) used in the mathematical implementation of the dual-source approach is mathematically equivalent to the logistic regression model (cf. section 2.2.1), then it should be possible to derive the same predictions from it. And it turns out that it is and that these predictions can also be derived from Bayes' theorem.[85] Moreover, they arguably also arise for the INUS theory.[86]

I suspect that the reason why this has apparently not been commented on in the literature is that the current practice is to estimate the values of the free parameters that achieve the most optimal fit without employing the qualitative constraints on the values estimates introduced here. So one way to make the problem go away is to continue as previously and merely to seek the optimal fit of the model to the data without worrying about how the values of the estimated parameters should be qualitatively constrained by the experimental conditions. However, this is obviously not very satisfying.

Another approach that I prefer is to view the models as describing the rational response and to regard deviations as indicators of irrationality on the part of the participants. The reason why I tend to prefer this approach is that both the influence of disablers on AC and of alternative antecedents on MT seem eminently reasonable, once one starts to think about it. If in the former case a predictor relationship between A and C is used to make the inference from the truth of C back to the truth of A, then presumably the endorsement rate of this inference should be weakened by disablers that target this predictor relationship. To illustrate using the special case of causal relations: if the causal relation between A and C is used to infer the occurrence of the cause based on the occurrence of the effect, then presumably disablers that weaken this causal relation should also have an impact on such abductive inferences. So it might, for instance, be reasonable to view intercourse with her boyfriend as a possible cause of Sally's pregnancy. But if told that he usually wears a condom, then the possibility of alternative causes should be assigned a greater weight than it was before.

To illustrate using Venn diagrams:

Figure 2: Venn diagram illustrating the AC disabler effect



We now see on the picture to the right that after more disablers have been added, the proportion of the Y=1 event taken up by the X=1 event has shrunken and the possibility of alternative antecedents as the cause of the Y=1 event should now be attributed a greater weight than before.

Of course, this effect only occurs, if Y=1 is not a subset of X=1. That is, the effect disappears, if there are no alternative antecedents and X=1 has a perfect degree of necessity for Y=1, where P(X=0, Y=1) = 0.

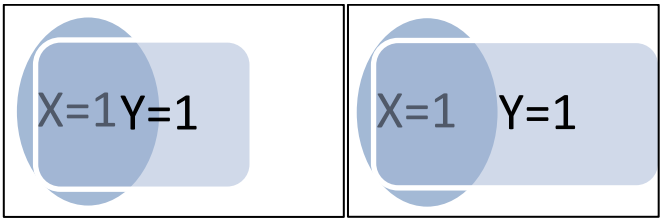
Turning to the influence of alternative antecedents on MT, we begin by noticing that there is no such influence in the case the antecedent has a perfect degree of sufficiency for the consequent where P(X=1, Y=0) = 0. Here both the logistic regression model and Bayes' theorem correctly assign a probability of one to MT:

$$MT = \frac{1}{1 + e^{ln\left(\frac{P(X=1,Y=0)}{P(X=0,Y=0)}\right)}} = \frac{1}{1 + e^{ln\left(\frac{0}{P(X=0,Y=0)}\right)}} = \frac{1}{1+0} = 1$$

$$P(X = 0|Y = 0) = \frac{P(X = 0, Y = 0)}{P(X = 0, Y = 0) + P(X = 1, Y = 0)} = \frac{P(X = 0, Y = 0)}{P(X = 0, Y = 0) + 0} = 1$$

So the influence of alternative antecedents on MT only occurs, whenever P(X=1, Y=0) > 0, and the antecedent has a less than perfect degree of sufficiency for the consequent. To illustrate the effect, consider the following Venn diagrams:

Figure 3: Venn diagram illustrating the MT alternative antecedent effect



What happens when we add more alternative antecedents to the picture is that the Y=1 event grows and the Y=0 event shrinks, as shown on the picture to the right. As a result, the X=1 ∩ Y=0 event now takes up a larger portion of the Y=0 event, and P(X=0|Y=0) has become less likely than before.

However, due to the difficulty of processing negations, it is harder to get one's head around this prediction than the effect of disablers on AC. So to ease the processing demands, lexicalized negations (e.g. losing) can be used as a substitute of explicit negations (not winning).[87] Let's suppose then that Y=1 is the event that the blue (underwaterrugby) team wins and that X=1 is the event that the blue captain is present. Then the conditional rule under consideration is 'if the blue captain is present, the blue team will win'. As the effect only holds for less than perfect degrees of sufficiency, it is required that a disabler like 'the blue captain is present but distracted by an important, upcoming exam' is active some of the time. This gives us the following inference:

$$\frac{\begin{array}{c}\textit{If the blue captain is present, blue team will win}\\ \textit{The blue team has lost}\end{array}}{\therefore \textit{The blue captain was absent}}$$

Now the point is that this inference should seem more likely before further alternative antecedents are added to the picture like that the blue team has now acquired the star player, Zack, who is fully capable of securing a victory, as a replacement for the blue captain, when he is absent.

The reason why the conclusion should now be viewed as less likely than it was before is that explaining the loss by the absence of the blue captain now also requires that he was not replaced by Zack. That is, it is now rarer that the blue team loses and the proportion of the cases, where the loss is due to the absence of the blue captain has now shrunken. As a

result, the possibility that the loss was due to the disabler that the blue captain was present but distracted by an important, upcoming exam (and therefore played terribly) should now be given a greater weight than it was before.

To some extent, it is not surprising that an effect of disablers for AC goes hand in hand with an effect of alternative antecedents for MT. The reason is that an AC inference with the conditional rule 'if p, q' has the same minor premise and conclusion as an MT inference with 'if ~p, ~q', and what is a disabler for 'if p, q' counts as an alternative antecedent for 'if ~p, ~q'. Furthermore, an AC inference for 'if ~p, ~q' has the same minor premise and conclusion as an MT inference for 'if p, q', and what is a disabler for 'if ~p, ~q' counts as an alternative antecedent for 'if p, q'. Finally, we notice the correspondence of the alternative antecedent effect on MT disappearing, if there are no disablers and the antecedent has a perfect degree of sufficiency for the consequent, and of the disabler effect on AC disappearing, if there are no alternative antecedents and the antecedent has a perfect degree of necessity for the consequent. So obviously, the two effects are closely related. However, the difficulty with processing negations makes the effect of alternative antecedents on MT less obvious than the effect of disablers on AC.

Given the considerations above, it seems rational that disablers should have an effect on AC and alternative antecedents should have an effect on MT. Yet, apparently the participants are not sensitive to such considerations. Hence, the failure of this prediction does constitute an inadequacy of the models in question when considered as descriptive models. However, the considerations above suggest that we should retain the models as normative models and attribute a failure of reasoning to the participants in the case of non-compliance. At this point it is important to be clear on that the point of this attribution is *not* to exonerate the models from their descriptive inadequacy.

Although one would ideally like a meta-analysis as a substitute for Politzer and Bonnefon's somewhat vague assessment,[88] producing these predictions constitutes a descriptive failure and attributions of failures of reasoning are not going to save the models from this predicament. Hence, if producing these predictions had only been a peculiarity of the logistic regression model and we couldn't make intuitive sense of them, we should probably leave it at that. But given that these predictions can also be derived from Bayes'

theorem and that intuitive examples can be formulated, the predictions become interesting as they now present us with an unsolved puzzle.

The utility of the present rationality assessment for empirical research thus consists in the new research questions it opens up for. For instance, inquiry can now be made of whether there are heuristics or biases that kick in under these circumstances and under what conditions the performance of the participants can be improved.

Moreover, given that it constitutes *a failure* not to comply with these predictions, questions can now be raised about whether it is connected to other known failures in judgments and decision making.

Interestingly, the failure to take the influence of disablers into account in one's abductive reasoning might be connected to a general confirmation bias. That is, this failing might be related to a general tendency to seek confirmatory evidence of one's own beliefs, which makes us less likely to take alternative hypotheses into account and evaluate counterevidence to the appropriate degree (cf. Nickerson, 1998). As the influence of the disablers weakens the causal relation, the subject should be less certain about seeing the occurrence of the effect as in agreement with his favorite causal hypothesis and should instead proceed to consider the degree to which it can be explained by alternative hypotheses. But if the subject has a general tendency to seek confirmatory evidence and ignore alternative hypotheses, he will fail to make such adjustments to his abductive reasoning in the face of disablers.

Similarly, the lack of sensitivity to the influence of alternative antecedents on MT might be connected to the confirmation bias. In a situation where the participants accept the conditional rule, they fail to adjust their expectations about exceptions to the conditional rule to the appropriate degree in the presence of alternative hypotheses. Again we see these two components at work: a failure to properly integrate the influence of alternative hypotheses into one's reasoning and a failure to adopt the right attitude towards the possibility of counterevidence to a conditional rule that one accepts.

However, due to the non-obviousness of the MT prediction and the difficulty of processing negations, a perhaps more plausible explanation is that the subjects simply fail to realize its correctness. Indeed, as neither the AC nor the MT prediction are commented

on in the literature (to the best of my knowledge), it is also possible that both are unobvious, normative implications that more or less go on unnoticed and that this fact explains the participant's lack of sensitivity to them.

So if one wants to test, whether the predictions have the envisaged connection to the confirmation bias, one possibility is to measure whether manipulations of the acceptability of the conditional rule has an influence on the lack of conformity to the effects in question. Perhaps the manipulation shouldn't take the form of obviously false conditional rules, as in Markovits & Shroyens (2007), because then the participants might be reluctant to draw the AC and MT inferences at all. But it could take the form of conditional rules that the participants are more or less neutral about. If lack of acceptance of the conditional rule has no influence on the conformity to the AC and MT predictions, then the latter are not specifically tied to a confirmation bias.

Notice finally that if the neglect of taking these adjustments to AC and MT into account had only been found in the overt experimental paradigm, where extra information about disablers and alternative antecedents is supplied, then the defect could also just be due to a failure to make appropriate adjustments to ensure that a probability distribution is produced after the extra information has been integrated. But as the overt and the covert experimental paradigms seem to produce the same finding, this explanation appears to be ruled out by the data.

4. The Logistic Regression Model and Fast & Frugal Heuristics

Before closing this chapter, we will briefly look at a possible objection to the logistic regression model, which derives from the work of Gigerenzer and his colleagues.

Gigerenzer, G., Todd, P. M., & the ABC Research Group (1999) have argued in general that an ecological and bounded approach to rationality should be adopted, which focuses on the adaption of cognitive mechanisms to the specific environments for which they were selected. This perspective is part of a general critical stance towards Kahneman and Tversky's famous heuristic and biases framework (see Kahneman, 2012). In contrast to

the latter, the research program of fast and frugal heuristics argues that the shortcuts that our cognitive system makes, which violate classical theories of rationality such as the probability calculus, need not be considered signs of irrationality, when the environmental constraints for which these shortcuts were adapted are taken into account. Indeed, Gigerenzer and his colleagues have made it a virtue to argue that simple heuristics may in many cases outperform the more advanced formalisms used in idealized theories of rationality.

One case in point is a heuristic called *take the best*, that we will learn more about below, which Gigerenzer and his colleagues have argued outperforms models of judgments based on multiple linear regression. Now since chapter III also drew an analogy between cognition and regression models, it is only appropriate, if we here consider a way of replying to Gigerenzer's criticism as it would apply to our logistic regression model.

It turns out that there is a long tradition for using multiple linear regression as a model of integration of cue information in predictions in the social judgment literature in relation to Brunswik's lens-model (Brehmer, 1994). However, this literature has not been connected to the literature on conditionals in psychology of reasoning.

A meta-analysis of five decades of research obtained from 86 articles shows that there is considerable evidence that subjects integrate information about multiple cues in a linear, additive fashion (Karelaia & Hogarth, 2008). One of the findings is that on average three statistically significant predictors are used by the subjects in such tasks. Generally, the regression models fit the data well. However, there is a worrying lack of consistency in the subjects' assignment of the regression weights, which increases with the complexity of the task and depends on how well the cues can be used to predict the values of the dependent variable in the specific task (Brehmer, 1994).

More recently, fast and frugal heuristics that make decisions based on one-reason decision rules have been advanced as an alternative to the regression models in modeling the performance on judgment tasks, where the participant has to decide which of two alternatives has the higher value on a numerical criterion given their values on a number of cues. An example is the *take the best* heuristic, which bases its decision on which of the two items meets the criterion in question (e.g. which of two professors has the highest income)

on the predictor with the highest weight that discriminates between the two, and merely guesses, if no such predictor can be found. In direct comparisons, this heuristic has been able to perform at the level of the regression model and sometimes even better, and there is accordingly ongoing research on under which environmental conditions heuristics that ignore information can be used with advantage (Gigerenzer, Todd, & the ABC Research Group, 1999: ch. 4-6, Gigerenzer & Brighton, 2009, Bröder, 2012).

However, the early studies suffered from methodological problems such as that all possible predictors (e.g. 8 or 15) were integrated into the regression equations, instead of only taking the statistically significant ones. Moreover, heuristics with dichotomous dependent variables were compared with multiple linear regression requiring interval scaled dependent variables, instead of using logistic regression.

In the only four studies that I know of that have implemented these requirements, the average number of predictors in judgment tasks for both the heuristic and the logistic regression turned out to be on the magnitude of 2 or 3 with the difference in several cases not being statistically significant (see Dhami & Harries (2001), Kee, Jenkins, McIlwaine, Patterson, Harper, & Shields (2003), Smith & Gilhooly (2006), and Backlund, Bring, Skånér, Strender, & Montgomery (2009)).

One thing to note, though, is that the dispute at stake here is over models of the cognitive *processes* underlying inference tasks, whereas the logistic regression model was a hypothesis about the underlying *knowledge representation*, which is an issue that this debate has been silent on. In principle, it is possible to exploit a knowledge representation that takes the form of a regression equation by means of the take the best heuristic by evaluating the weights on the basis of relative frequencies, and basing the decision on the predictor with the highest weight that can discriminate between the two items, instead of computing the answer by taking all the predictors into account. Which strategy is adopted might depend on such factors as time pressure, cognitive load, and motivation.[89]

Moreover, it is worth noticing that Gigerenzer's criticism did not concern use of regression models for *conditional reasoning*, which was the focus of our logistic regression model in chapter III. And indeed no heuristic alternative has yet been formulated for conditional reasoning that I am aware of (although such alternatives are in the making).[90]

We may finally also notice that the danger of overfitting does not arise for our logistic regression model in the way that Gigerenzer envisages, because the equations from chapter III only incorporate one predictor and not 8 or 15 as in some of his examples. However, this danger of overfitting is one of the main complaints that Gigerenzer and his colleagues have against the use of regression models. Hence, I conclude that his criticism against use of regression models in psychology of judgments does not straightforwardly carry over to the present usage in conditional reasoning.

If in the future heuristic alternatives are formulated for conditional reasoning, then it will, of course, be interesting to compare their performance to the logistic regression model advanced here. But we will not get the opposition between regression models overfitting by using 8 or 15 cues and fast and frugal heuristics using only one cue that Gigerenzer and his colleagues have set up in psychology of judgments as the logistic regression model only incorporates one predictor explicitly. (Of course, implicitly the presence of alternative antecedents influences the parameters of the model, as we have seen. But then again it is hard to see how one could possibly accommodate the results from the overt and covert experimental paradigm from section 3 without making such an assumption.)

Appendix 3:
On Learning Conditional Information

In Douven (2012) it is argued that strict Bayesian accounts have a problem with accounting for how we can update on information received in a conditional form. As it is pointed out, one difficulty in providing such an account is that the jury is still out on whether conditionals express propositions, yet the conditionalization procedure requires a propositional input.

As it turns out, section 2.2 already contained a strategy for dealing with this problem, which has the nice property of being neutral on the controversial issue of whether indicative conditionals express propositions. The general idea was that the hearer forms an expectation that a predictor relationship is being expressed by the use of indicative conditionals, which is emphasized through use of the predictive modal 'will'. As a result, what the hearer must do when learning conditional information is to set the regression weights of the assumed predictor relationship to default values, which are subject to modification through linguistic qualifiers. More specifically, in section 2.2.2 it was argued in the context of modeling the presence of the rule in the conditional inference task that expectation of both the perceived sufficiency and relevance of A for C should increase (whenever the 'circumstances are such that' reading of the conditional can be presupposed). This interpretation produced the prediction that MP↑↑, MT↑, and AC↑ in the presence of the rule.

Keeping these assumptions in mind, we now turn to three test cases that Douven (2012: 241) uses as criteria of adequacy on any account of what is learned through conditional information:

> Example 1
>
> Sarah and her sister Marian have arranged to go for sundowners at the Westcliff hotel tomorrow. Sarah feels there is some chance that it will rain, but thinks they can always enjoy the view from inside. To make sure, Marian consults the staff at the Westcliff hotel

and finds out that in the event of rain, the inside area will be occupied by a wedding party. So she tells Sarah:

(1) If it rains tomorrow, we cannot have sundowners at the Westcliff.

Upon learning this conditional, Sarah sets her probability for sundowners *and* rain to 0, but she does not adapt her probability for rain.

Example 2

Harry sees his friend Sue buying a skiing outfit. This surprises him a bit, because he did not know of any plans of hers to go on a skiing trip. He knows that she recently had an important exam and thinks it unlikely that she passed. Then he meets Tom, his best friend and also a friend of Sue's, who is just on his way to Sue to hear whether she passed the exam, and who tells him:

(2) If Sue passed the exam, her father will take her on a skiing vacation.

Recalling his earlier observation, Harry now comes to find it more likely that Sue passed the exam.

Example 3

Betty knows that Kevin, the son of her neighbors, was to take his driving test yesterday. She has no idea whether or not Kevin is a good driver; she deems it about as likely as not that Kevin passed the test. Betty notices that her neighbors have started to spade their garden. Then her mother, who is friends with Kevin's parents, calls her and tells her the following:

(3) If Kevin passed the driving test, his parents will throw a garden party.

Betty figures that, given the spading that has just begun, it is doubtful (even if not wholly excluded) that a party can be held in the garden of Kevin's parents in the near future. As a result, Betty lowers her degree of belief for Kevin's having passed the driving test.

I take it that a reasonable interpretation of these examples runs as follows. Example 1 illustrates that P(A) should not be affected by the addition of conditional information in the absence of any further information.

The strategy outlined above is capable of accommodating this effect. The reason is that the increase in perceived sufficiency introduced by the conditional rule has been interpreted in section 2.2.2 as having the effect of decreasing P(X=1, Y=0), which is counterbalanced by an increase to P(X=1, Y=1) to ensure that we still end up with a probability measure. In example 1, we have a negation in the consequent, so here the effect is reversed with P(X=1, Y=0)↑ and P(X=1, Y=1)↓. But the result is the same; to wit, that the changes are counterbalanced so that P(X=1) remains unaffected in the absence of further information. And this gives us the result we need for dealing with the first example.

Example 2 illustrates an increase in P(A) due to use of AC upon both learning the conditional rule and that P(C) = high, and example 3 illustrates a decrease in P(A) due to use of MT upon both learning the conditional rule and that P(C) = low.

As we noted, the presence of the conditional rule increases the perceived sufficiency by decreasing P(X=1, Y=0) and increasing P(X=1, Y=1), which in turn will have the effect of MP↑↑, MT↑, and AC↑. Hence, the conditional rule is predicted to increase P(X=1) in example 2, where the additional information of P(Y=1) = high is supplied, as the agent can now use his auxiliary information that P(Y=1) = high to draw an AC inference, and as the presence of the conditional rule sanctions an increase in the endorsement rate of the latter type of inference. Hence, example 2 appears to be compatible with the predicted effect of the presence of the conditional rule.

Moreover, the conditional rule is predicted to decrease P(X=1) in example 3, where the additional information of P(Y=0) = high is supplied, as the agent can now use his auxiliary information that P(Y=0) = high to draw an MT inference, and as the presence of the conditional rule sanctions an increase in the endorsement rate of the latter type of inference. Hence, example 3 appears to be compatible with the predicted effect of the presence of the conditional rule.

We can thus conclude that the present model is capable of performing satisfactory on Douven's (2012) three test cases.

[74] Acknowledgment: this chapter profited from discussions with and comments by Sieghard Beller, Wolfgang Spohn, Karl Christoph Klauer, Henrik Singmann, and Igor Douven.

[75] Explication of deductive instructions: I.e. instructions stressing the irrelevance of background knowledge and necessary truth preservation, which use a binary response format (or a ternary – if 'don't know' is included as an option).

[76] Explication of inductive instructions: I.e. instructions stressing the degree to which they accept the conclusion on the basis of background knowledge, which uses a graded response format.

[77] Terminological warning: care must be taken, because 'C' has been used to represent the consequent in conditionals throughout the dissertation, and now it is being used in equation (2) to represent the particular content of an inference problem.

[78] Philosophical implications: in itself this aspect of Spohn's (2012: ch. 6) explication of reasons has profound implications for, where to draw the boundary between epistemic agents, such as humans, and animal cognitive systems. In philosophy there is a tradition culminating in Brandom (1994) of drawing the line of demarcation at the understanding of reasons. But if the reason relation is already involved in classical conditioning as we saw in chapter III, then this cannot be quite right. Perhaps the boundary must then be drawn between use of the reason relation in sheer prediction and use of the reason relation in argumentative contexts, where justificatory challenges are made and epistemic responsibility becomes an issue.

[79] Reference: Klauer *et al.* (2010).

[80] Acknowledgement: thanks to Karl Christoph Klauer for pressing me on this issue.

[81] Exception: however, one exception with respect to sufficient and insufficient reasons is Neth & Beller (1999).

[82] General effects of negation: in general it is known that negations are hard to process and it is believed that this might be part of the reason why the participants have their difficulties with MT inferences. However, one important exception to this rule of inferior performance with negations is introducing negations in the consequent of the conditional rule in the Wason selection task (cf. endnote 109). The reason is that the participants appear to restrict their reasoning to the cards mentioned in the conditional rule regardless of the polarity due to a simple matching bias. So when the conditional rule takes the form of 'if *p*, *q*', then the participants will tend to make the incorrect *p* & *q* selection as opposed to the correct *p* & ~*q* selection. If, on the other hand, the conditional rule is 'if *p*, ~*q*', then their matching bias will once more restrict their reasoning to the *p* & *q* cards—only this time it makes them select the right card combination (Evans & Over, 2004: ch. 5).

[83] Examples of use of auxiliary hypotheses : in psychology of reasoning, general working memory limitations and Gricean principles are probably the best examples of auxiliary hypotheses that are not

systematically related to the core claims of the theories in question, but nevertheless regularly invoked to increase the theories compatibility with existing findings. Moreover, in accounting for a minor effect in Klauer *et al.* (2010), we ourselves had to resort to the use of auxiliary hypotheses in section 2.2.2.

[84] Potential objection: as said, the predictions listed concern the rule-free baseline condition, whereas the Politzer & Bonnefon quote concerns the results in experimental paradigms, where the conditional rule is present. But this should not really make much of a difference as manipulations of disablers and alternative antecedents have similar effects under both conditions.

[85] Predictions: MP: $1 - e(C) = 1 - P(Y = 0|X = 1) = \frac{P(X=1,Y=1)}{P(X=1)} = \frac{P(X=1,Y=1)}{P(X=1,Y=1)+P(X=1,Y=0)}$

MT: $\frac{1-b(C)-a(C)\cdot e(C)}{1-b(C)} = \frac{\mathrm{P(Y=0)-P(X=1)\cdot P(Y=0|X=1)}}{\mathrm{P(Y=0)}} = \frac{P(Y=0,X=0)}{P(Y=0,X=0)+P(Y=0,X=1)}$

AC: $\frac{a(C)(1-e[C])}{b(C)} = \frac{\mathrm{P(X=1)\cdot P(Y=1|X=1)}}{\mathrm{P(Y=1)}} = \frac{P(X=1,Y=1)}{P(X=1,Y=1)+P(X=0,Y=1)}$

DA: $\frac{1-b(C)-a(C)\cdot e(C)}{1-a(C)} = \frac{\mathrm{P(Y=0)-P(X=1)\cdot P(Y=0|X=1)}}{\mathrm{P(X=0)}} = \frac{P(X=0,Y=0)}{P(X=0,Y=0)+P(X=0,Y=1)}$

Since the presence of disablers has the effect of P(X=1, Y=0)↑ and (X=1, Y=1)↓, its effect on the equations above is: MP↓↓, MT↓, AC↓. Since the presence of alternative antecedents has the effect of P(X=0, Y=1)↑ and (X=0, Y=0)↓, its effect on the equations above is: AC↓, DA↓↓, MT↓.

Unsurprisingly, one can arrive at the same equations by means of Bayes' theorem:

MP: $P(Y = 1|X = 1) = \frac{P(X=1,Y=1)}{P(X=1)} = \frac{P(X=1,Y=1)}{P(X=1,Y=1)+P(X=1,Y=0)}$

MT: $P(X = 0|Y = 0) = \frac{P(X=0,Y=0)}{P(Y=0)} = \frac{P(X=0,Y=0)}{P(X=0,Y=0)+P(X=1,Y=0)}$

AC: $P(X = 1|Y = 1) = \frac{P(X=1,Y=1)}{P(Y=1)} = \frac{P(X=1,Y=1)}{P(X=1,Y=1)+P(X=0,Y=1)}$

DA: $P(Y = 0|X = 0) = \frac{P(X=0,Y=0)}{P(X=0)} = \frac{P(X=0,Y=0)}{P(X=0,Y=0)+P(X=0,Y=1)}$

Hence, the predictions in question can also be derived from it.

[86] The predictions and the INUS theory: a case can be made that the INUS theory advanced in Klauer (manuscript) is committed to making these predictions as well due to the causal interpretation of these predictions given in the text below. The reason is that the theory holds that AC and MT inferences are given the following causal reading: "AC: How likely did AX rather than an alternative cause Y cause B, given B? MT: How likely did the causal path from A to B not come into being, given the absence of the effect, due to A not occurring rather than due to X not being in force?" (p. 8). [Here 'X' denotes an enabling condition of A.]

[87] Reference: cf. Sperber, Cara, & Girotto (1995).

[88] An exception: for instance, most of the results in Neth & Beller (1999) and in Klauer *et al.* (2010) constitute exceptions, where disablers influence AC and DA and alternative antecedents influence MP and MT.

[89] Explication of a surprising empirical finding: however, somewhat surprisingly, Bröder (2012) found that 73 % percents of the participants were able to use the allegedly more demanding linear integration of all the predictors strategy even under heavy cognitive load, and it turned out that although the simple heuristics are able to perform well under specific environmental conditions, they introduce the cost of devoting cognitive resources to selecting the proper strategy. Furthermore, although one would expect the heuristics to be the less sophisticated alternative, it has been found that novices tend to rely on a weighted linear integration of information (Garcia-Retamero & Dhami, 2009). Moreover, it turns out that in novel laboratory tasks the majority starts out with the latter strategy, and that it is actually the group with the higher IQ scores that are able to change to the heuristics, when incentives and environmental structures are introduced that favor TTV (Bröder, 2012). Hence, these findings put the common impression into perspective that fast and frugal heuristics are supposed to be less demanding.

[90] Reference: Keith Stenning and Laura Martignon (personal communication).

V

Logical Omniscience and Acknowledged vs. Consequential Commitments[91]

Abstract: With chapter IV, our attempt of implementing the methodological recommendations of chapter I with respect to the ranking-theoretic approach to conditionals has come to an end. It is now time to shift gears for a second time and to consider what to do about the idealizing assumptions of ranking theory in light of the recent rationality debates in cognitive psychology. More specifically, it is to be investigated what explanatory resources Robert Brandom's distinction between acknowledged and consequential commitments affords in relation to the problem of logical omniscience. With this distinction the importance of the doxastic perspective under consideration for the relationship between logic and norms of reasoning is emphasized, and it becomes possible to handle a number of problematic cases discussed in the literature without thereby incurring a commitment to revisionism about logic. As we shall see, the problem of logical omniscience not only arises within ranking theory but also within the recent paradigm shift in psychology of reasoning. So dealing with this problem is important not only for philosophical purposes but also from a psychological perspective.

1. Introduction

Like other models in formal epistemology, ranking theory is based on the following norms: (i) rational beliefs are deductively closed, (ii) rational beliefs are completely consistent, and (iii) every logically equivalent sentence is always believed to the same degree by the rational agent (Spohn, 2012: ch. 4-5, Huber, 2013). However, as Spohn (2012: 79, 2013b) points out, it can be shown on the basis of an axiomatization of ranking theory in terms of conditional negative ranking functions that the norm of conditional consistency already entails the deductive closure of rational beliefs in ranking theory. So there is a sense in which this norm is the most basic in ranking theory.

Now if ranking theory is to have any applications to psychology of reasoning, it is useful to step back from the detailed discussion of conditionals that we have been conducting to take a synoptic view and consider the question of whether the normative foundation of ranking theory is too idealized to be applicable to real agents. The way that the present chapter deals with this issue is by presenting one strategy for making it less idealized. It does this by considering the explanatory resources that Brandom's (1994) distinction between acknowledged and consequential commitments affords in relation to the problem of logical omniscience.[92] Hence, one of its goals is to use existing literature to identify a number of problems that any adequate account of the relation between norms of reasoning and logic should be capable of meeting (section 3). In a second step it will then be shown how a particular approach based on the abovementioned conceptual distinction is capable of delivering (what appears to be) satisfactory answers to all of them (section 4).

Briefly stated, *the problem of logical omniscience* is the problem that (i)-(iii) appear to impose too demanding constraints on real agents (cf. Stalnaker, 1999: ch. 13-14, Levi 1991: ch. 2, 1997: ch. 1). So whereas we have the abovementioned move away from theories based on deductive logic in psychology of reasoning due to the poor logical performance documented in the psychological literature (Evans 2002, 2012, Oaksford & Chater, 2007, 2010), it is customary to treat deductive closure and consistency as minimal conditions for belief sets in formal epistemology. And if the objects of beliefs are taken to be propositions, then logically equivalent sentences are automatically treated as being believed

to the same degree—irrespectively of well-known psychological findings such as the framing effect (Kahneman, 2012).[93]

In addition to such discrepancies with well-established empirical findings, these norms of rational belief have also come under considerable pressure from a range of problematic cases cited in the philosophical literature, which are introduced in section 3. So both the psychological and philosophical literatures suggest that the status of these minimal constraints on belief sets needs to be carefully scrutinized. However, it should be noted that the normative principles in question are as much a part of ranking theory, and related logic-based approaches like belief revision theory, as they are of the probabilistic models that psychology of reasoning has begun to import from Bayesian epistemology.

Christensen (2007: 15ff.) thus argues that the probability calculus should not be seen as a new logic for graded belief, but rather as "a way of applying standard logic to beliefs, when beliefs are seen as graded". He makes his case by showing on the basis of the axioms of the probability calculus how the logical properties of propositions impose restrictions on probabilistic coherence. An example is that probabilistic coherence requires of the agent that he believes $p \vee q$ at least as strongly as p, which follows directly from the fact that $p \vee q$ is entailed by p. Hence, just as logical closure for binary beliefs would require that the ideally rational agent does not believe p while not believing $p \vee q$, so probabilistic coherence for graded beliefs requires of him that he does not believe p to degree x while believing $p \vee q$ to a degree less than x. Moreover, just as logical consistency of binary beliefs would require that this agent doesn't believe both p and $\neg (p \vee q)$, probabilistic coherence of graded beliefs requires that his degree of belief in p and $\neg (p \vee q)$ does not sum up to more than one (Ibid.: 15-16).

So no matter whether binary, formal representations of beliefs are preferred (as in the old paradigm in psychology of reasoning), or probabilistic representations of degrees of beliefs are preferred (as in the new paradigm in psychology of reasoning), it holds that: "the prominent proposals for imposing formal constraints on ideal rationality are rooted in logic" (Ibid.: 18). It is only recently that there has been an awareness of this fact in the psychological literature. Evans (2012: 6) has aptly put his finger on the implication this has for the celebrated paradigm shift in psychology of reasoning when he says:

> By around 2000 many researchers using the paradigm were questioning the idea that logic could provide a description of human reasoning, and many were also casting doubt on logic as an appropriate normative system (Evans, 2002; Oaksford & Chater, 1998). While these authors complained about "logicism" in the psychology of reasoning, it is again standard bivalent logic that they had in mind. Any well-formed mathematical system is a closed deductive system that can be regarded as a logic in which theorems (proven conclusions) are deduced from axioms (assumptions). Probability theory, which is much used in the new paradigm, actually reduces to binary logic when probabilities are set to 1 or 0. For example, if we set P(A and B) = 1, we can conclude that P(A) = 1, thus preserving certainty (truth). So it is more accurate to say that authors were objecting to binary logic, which does not allow beliefs represented as subjective probabilities that range freely from 0 to 1, rather than logic per se.

Accordingly, the shift in psychology of reasoning is to be viewed as one concerning the need for representing degrees of beliefs that are concerned with our confidence in propositions rather than with necessary truth preservation of full beliefs. Yet, because the minimal constraints on belief sets have not been abandoned, we are still confronted with the problem of logical omniscience.

In this context, Brandom (1994) has made an interesting conceptual distinction between acknowledged and consequential commitments, which can potentially throw new light on the normative issues at stake. Section 2 therefore introduces the pertinent features of his account.

2. Acknowledged and Consequential Commitments

2.1 Introducing the Brandomian Framework

Instead of theorizing about belief, Brandom (1994) chooses to theorize about public, doxastic *commitments*, which conversation partners attribute to one another on the basis of the assertions they make and whether they later withdraw them. In this type of interaction, the interlocutors alternate between taking up the role of *the speaker*, who makes the

assertions, and *the scorekeeper*, who keeps track on the assertions made by the speaker by keeping score on the speaker's commitments and entitlements.

A doxastic commitment to *p* can be thought of as an obligation to defend *p* when appropriately challenged. For some of an agent's doxastic commitments it holds that the agent already counts as has having redeemed his obligation to defend the corresponding claims (either because there are no standing challenges to his warrant that cannot be met, or because the claims are so trivial that they *per default* have a defeasible status of not being in need of justification). For the commitments for which this holds, the agent is said to be (defeasibly) *entitled* to his assertions. Moreover, when a claim is attributed entitlement, it becomes possible for others to adopt a commitment to the claim in question while deferring back to the original speaker for the burden of justification.

To introduce the distinction between acknowledged and consequential commitments, Brandom says:

> The commitments one is disposed to avow are *acknowledged* commitments. But in virtue of their inferentially articulated conceptual contents, assertional commitments have consequences. Undertaking a commitment to a claim with one content involves undertaking commitments to claims whose contents are (in the context of one's other commitments) its committive-inferential consequences. Undertaking a commitment to the claim that Pittsburgh is to the West of Philadelphia is one way of undertaking commitment to the claim that Philadelphia is to the East of Pittsburgh. These *consequential* commitments may not be acknowledged; we do not always acknowledge commitment to all the consequences of the commitments we do acknowledge. They are commitments nevertheless. (1994: 194)

For some of the doxastic commitments undertaken by the speaker, the scorekeeper will in other words note that they are acknowledged by the speaker. For others the scorekeeper can note that they are consequences of the *acknowledged commitments*, which the speaker might not acknowledge. One way of thinking about the underlying issue is this: by making an assertion one adopts a conditional task responsibility to defend the claim in light of appropriate challenges. And if a doxastic commitment has other doxastic commitments as its consequences, then their falsity can be made part of the challenge posed to attempts

of justifying the original claim, even if the speaker is ignorant of the consequences of what he is saying. To take an example, suppose a speaker asserts both that 'Berlin is to the North of Behrendorf' and 'Copenhagen is to the South of Behrendorf', then the scorekeeper may challenge these claims by pointing out that they introduce a consequential commitment to the claim that 'Berlin is to the North of Copenhagen' due to transitivity, and that we know the latter claim to be mistaken.

But to connect the present considerations to the issue of deductive closure above, it must be observed that Brandom talks about consequential commitments in relation to *material* (committive) *inferences* like the inference from one location being west of a second location to the second being east of the first.[94] Nowhere does he raise the issue in relation to the logical consequences of one's beliefs that I am aware of. However, this shortcoming can easily be remedied, because Brandom analyzes the inferential articulation of conceptual content as consisting in the following relations (Brandom, 1994, MacFarlane, 2010):

> *Commitment preservation*: The inference from premises Γ to *q* is *commitment-preserving* if a commitment to Γ counts as a commitment to *q*.
>
> *Entitlement preservation*: The inference from premises Γ to *q* is *entitlement-preserving* if an entitlement to Γ counts (defeasibly) as an entitlement to *q*.
>
> *Incompatibility*: *p* is incompatible with *q* if a commitment to *p* precludes an entitlement to *q*.

Since Brandom says that commitment-preserving inferences generalize the category of *deductive inferences*, and entitlement-preserving inferences generalize the category of *inductive inferences*, it seems reasonable as a first approximation to explicate the underlying reason relations in terms of Spohn's (2012: ch. 6) account of reasons as follows:

> *Commitment preservation*:
> $\tau(q|\Gamma) > \tau(q|\Gamma^C), \tau(q|\Gamma) = \infty$[95]
> or
> $P(q|\Gamma) > P(q|\Gamma^C), P(q|\Gamma) = 1$

Entitlement preservation:

$\tau(q|\Gamma) > \tau(q|\Gamma^C)$, $\tau(\Gamma) > a$, $\tau(q|\Gamma) > a$, for $a \geq 0$

or

$P(q|\Gamma) > P(q|\Gamma^C)$, $P(\Gamma) > b$, $P(q|\Gamma) > b$, for $b \geq 0.5$[96]

where a and b denote a contextually set threshold of when the speaker counts as having fulfilled his obligation to defend his assertions.

Moreover, it is possible to formulate both a weak and a strong notion of incompatibility, where the latter is the limiting case of the former and the case of logical inconsistency is an instance of strong incompatibility:

Weak Incompatibility:

$\tau(q|p) < \tau(q|\neg p)$, $\tau(q|p) < a$, for $a \geq 0$

or

$P(q|p) < P(q|\neg p)$, $P(q|p) < b$, for $b \geq 0.5$

Strong Incompatibility:

$\tau(q|p) < \tau(q|\neg p)$, $\tau(q|p) = -\infty$

or

$P(q|p) < P(q|\neg p)$, $P(q|p) = 0$

Hence, what was said about consequential commitments above should *ipso facto* apply to the logical consequences of the speaker's doxastic commitments, and what Brandom says about incompatibility should *ipso facto* apply to the case of logical inconsistency, and we can thus begin to apply our conceptual distinctions to the problem of logical omniscience below in sections 3 and 4. (However, beyond this observation, the explications given above, which depict Brandom's inferential semantics as a ranking-theoretic (or probabilistic) reason relation semantics,[97] will play no further role in the course of the argument.)

The point of introducing the distinction between acknowledged and consequential commitments is to avoid an ambiguity in belief talk:

> In one sense, one believes just what one takes oneself to believe, what one is prepared to avow or assert. In another sense, one believes, willy-nilly, the consequences of one's beliefs (…). The sense of belief in which one is taken actually to believe what one ideally *ought* to believe (at least given what else one believes), call it *ideal* or *rational* belief, can conflict with the sense of belief for which avowal is authoritative. (…) The conflict arises precisely because one can avow incompatible beliefs, and fail to avow even obvious consequences of one's avowals. (Brandom, 1994: 195)

When we leave beliefs behind and focus on public, doxastic commitments, the analogue to cases of incompatible beliefs gets analyzed as cases, where *incompatible obligations* to defend claims have been undertaken. That is, such cases are viewed as the doxastic counterpart to cases, where agents have undertaken incompatible practical commitments by, for example, promising to be in two different places at once (Brandom, 1994: 196). In both cases we are dealing with instances of our general shortcoming as agents that we sometimes undertake multiple obligations that cannot all be redeemed at the same time.

Where things begin to become interesting is in relation to consequential commitments. As Kibble (2006b: 37) points out, just as it would be an inappropriate response to an agent, who has undertaken incompatible practical commitments, to attribute any arbitrary intention, it is a central feature of Brandom's pragmatic model of giving and asking for reasons that it would be inappropriate to follow the principle of *ex falso quodlibet* and attribute any arbitrary doxastic commitment to an agent, who has undertaken incompatible doxastic commitments. Instead the appropriate response is to withhold attributions of entitlement to the particular claims that are incompatible (Brandom, 1994: ch. 3). Through this act, any further inheritance is blocked to these claims through testimony that would have allowed other agents to adopt a commitment to them while deferring back to the speaker for the burden of justification. Yet, this need not commit us to revisionism about logic, as we shall see in section 4.

2.2 Reinterpreting the Norms of Rational Belief

It is worth noticing that—as Milne (2009: 276) points out—the principles of rationality have a natural justification on the basis of the norms of assertion. Extending a bit, the argument would go roughly as follows:

(P_1) Making an assertion is to be understood as licensing others to use it as an uncontroversial starting point for further inquiry[98] while deferring back to the speaker for the burden of justification (cf. Brandom 1994: 174, 2001:165).

(P_2) The interlocutors would not be able to use an inconsistent set of propositions as an uncontroversial starting point for further inquiry.

(P_3) The interlocutors would not be able to use a set of propositions that have unacceptable logical consequences as an uncontroversial starting point for further inquiry.

(P_4) The interlocutors would not be able to use the speaker's assertions as an uncontroversial starting point for further inquiry, if they have logically equivalent formulations that are themselves unacceptable.

(C) Hence, the speaker's obligation to defend the assertions he makes when appropriately challenged extends to avoidance of their inconsistency, defending their logical consequences, and to defending their logically equivalent formulations.

Essentially the idea is that it is part of the epistemic use to which the speaker's interlocutors can reasonably put his assertions to exploit their logical properties for further computation, which means that it would constitute a failure, if the speaker feeds them assertions that fail to meet its minimum requirements. As a result, the speaker's interlocutors are entitled to enlist the logical consequences of his acknowledged commitments as consequential commitments with an equal claim to form the basis of challenges as his acknowledged commitments.

Following this line, we can begin to view the minimal rationality constraints on beliefs sets introduced in section 1 as constraints governing the score of commitments and

entitlements that the scorekeeper keeps on the speaker in the course of an argumentative dialogue. That is, in deciding whether the speaker has a constellation of commitments for which it both holds that there are no serious, unmet justificatory challenges, and that others would be permitted to inherit claims while deferring back to the speaker for the burden of justification, the scorekeeper can be seen as being engaged in the task of constructing a belief set based on the speaker's public utterances that is to be consistent and closed under logical consequence.

Viewing matters from this perspective allows us to regard the importance of these rationality principles as not consisting in whether speakers actually succeed in only avowing to consistent beliefs and all their logical consequences (which would be a claim of which the empirical literature suggests that we should be highly skeptical). But rather as consisting in there being norms that we impose on others, when deciding whether it is safe to accept what they say, which we hold them accountable to in justificatory challenges. That is, what matters in this context is not so much the speakers' actual performance in their own individual reasoning, but whether they would accept challenges of their claims based on: (1) documented inconsistencies, (2) logical consequences of their claims that are themselves unacceptable, and (3) logically equivalent formulations of their claims that are themselves unacceptable. If the speakers accept such challenges, they can be taken to display the recognition of being bound by these norms even if they are unable to comply with them by their own efforts.

3. Four Possible Gaps between Logic and Norms of Reasoning

In an unpublished manuscript that is too good not to be cited, MacFarlane (manuscript) considers 36 possible bridge principles between norms of reasoning and logical consequence that take the following form:[99]

If A, B, ⊨ C, then (normative claim about believing A, B, and C)

The different versions are produced by varying the following four parameters: (1) the type of deontic operator (i.e. whether facts of logical validity give rise to *obligations*, *permissions*, or *defeasible reasons* for beliefs), (2) the polarity (i.e. whether the obligations, permissions, or defeasible reasons concern *believing* or *not disbelieving*), (3) the scope of the deontic operator, and (4) whether the facts about logical validity have to be known by the agent. But the preceding discussion has already brought out further parameters that MacFarlane's otherwise comprehensive discussion fails to consider: (5) beliefs vs. public commitments, (6) acknowledged commitments vs. consequential commitments, and (7) the doxastic perspective of the speaker vs. that of the scorekeeper.

So to illustrate the attractiveness of transposing the normative issues in the way outlined above by thinking of the rationality principles not as principles of private beliefs, but as principles of public commitments, which are imposed from a scorekeeping perspective, it is instructive to review some of the puzzle cases that MacFarlane discusses. More specifically, we are going to look at the arguments posed by Harman (1986) to show the lack of a connection between logical consequence and norms of reasoning, which have been succinctly summarized by Hartry Field (2009: 252-3) as follows:

> Problem 1:
> Reasoning (change of view) doesn't follow the pattern of logical consequence. When one has beliefs A_1, …, A_n, and realizes that they together entail B, sometimes the best thing to do isn't to believe B but to drop one of the beliefs A_1, …, A_n.
>
> Problem 2:
> We shouldn't clutter up our minds with irrelevancies, but we'd have to if whenever we believed A and recognized that B was a consequence of it we believed B.
>
> Problem 3:
> It is sometimes rational to have beliefs even while knowing they are jointly inconsistent, if one doesn't know how the inconsistency should be avoided.

Problem 4:

> No one can recognize all the consequences of his or her beliefs. Because of this, it is absurd to demand that one's beliefs be closed under consequence. For similar reasons, one can't always recognize inconsistencies in one's beliefs, so even putting aside point 3 it is absurd to demand that one's beliefs be consistent.

An example of problem 3 is the preface paradox, where the author of a book finds that he has supporting evidence for every single claim made in his book, yet knowledge of his own general fallibility cautions him to disbelieve the conjunction of all his claims. If beliefs are closed under conjunction, he thereby finds himself with an inconsistent belief set. Yet it is not clear what he should do about it as all of his beliefs seem quite reasonable.

A further example is given by Hartry Field in his second John Locke lecture:

> any rational person would have believed it impossible to construct a continuous function mapping the unit interval onto the unit square until Peano came up with a famous proof about how to do it, so the belief that no such function could exist was eminently rational but inconsistent, and there are many more examples of a similar nature.[100]

Below in section 4 bridge principles will be formulated that differ from those MacFarlane considers by introducing parameters (5)-(7), which are capable of handling problems 1-4 as well as three further constraints that MacFarlane (manuscript) considers. This is a significant contribution as MacFarlane presents these desiderata as standing in a tension and thus requiring some sort of trade-off. But first we start out with some initial observations and a treatment of the preface paradox.

3.1 Preliminary Observations

The first thing to notice is that we can simply grant Harman (1986), Foley (1993), and others that there are cases like the preface paradox, where it, from the speaker's point of view, may make sense to give in and learn how to live with an inconsistency, if it is either too hard or costly to deal with the problem. Moreover, logic does not provide a guide for the speaker for how to manage his acknowledged commitments, if it comes to his attention

that they have a logical consequence that is better avoided, because there are always more ways of resolving the issue as problem 1 indicates.

Yet, this does not mean that the principles of rationality cease to impose norms of reasoning, and that the scorekeeper should cease to treat the speaker as *obligated* to avoid inconsistencies and accept the logical consequences of his acknowledged commitments (as long as they have not been withdrawn) as we shall see in detail below. Furthermore, the speaker can be seen as recognizing that these norms are still in force, if he accepts the appropriateness of challenges based on his failure to repair his "public belief set".

As we have seen, the outcome of the scorekeepers' failure to construct a deontic score for the speaker that meets the minimal constraints on belief sets is not that the speaker fails to have any rational beliefs. For first of all, we are treating these principles as requirements of public commitments and not as requirements of (rational) beliefs. Secondly, the speaker's failure to comply with them does not even mean that he does not have any public, doxastic commitments. It just means that he has undertaken an obligation to defend a constellation of claims that he cannot redeem (either because they are directly inconsistent, because they have logically equivalent formulations that cannot be defended, or because they would require him to accept as consequential commitments logical consequences of his claims, which in turn cannot be defended). Thirdly, the consequence of this failure is that the speaker for the moment cannot be attributed entitlement (and be treated as a source of entitlement for others) *with respect to the afflicted assertions.*[101] But this may be a consequence that the speaker may have to live with at times, where there is no obvious repair to the constellation of obligations that he has undertaken. The rationale for this penalty is to avoid the propagation of error, and indeed both Foley (1993: 119) and Harman (1986: 15-7) agree that it would be a mistake to base further inquiry on inconsistent propositions even if they are sometimes unavoidable.

Because the consequential commitments are only used as an aid in deciding, whether entitlement can be attributed, the possibility is not precluded that the speaker may sometimes be rationally permitted to manage his acknowledged commitments in ways that temporarily exclude him from attributions of entitlements. In such cases, the agent's assertions can be treated temporarily as not being a source of information that can be used

unproblematically as a base for further inquiry. If it happens regularly, then the agent can be blacklisted (see also Kibble (2006b)). In this way it is possible to drive a wedge between our assessments of the agent's rationality and of the information that we want to use for further inquiry. For rational agents it need not be possible to be a source of valuable information under all circumstances—no matter how paradoxical the requirements they are confronted with.

A case in point may be the preface paradox, which we will return to shortly. In this context, it is also worth noting the situation that Harman (1986: 16) argues that most of us are in when it comes to the liar paradox:[102]

> the rational response for most of us may simply be to recognize our beliefs about truth are logically inconsistent, agree this is undesirable, and try not to exploit this inconsistency in our inferences.

Furthermore, Foley (1993: 115-7) discusses a number of interesting cases, where he, *inter alia*, makes the point that sometimes the optimal strategy is not the one that has a small chance of arriving at an ideal outcome, where no mistakes are made, but rather one that minimizes the expected number of mistakes (even if one can thereby be certain that mistakes are made some of the time). Indeed, a case could be made that this is exactly the type of situation we find ourselves in, when we have to rely on what is known to be fallible sources of information, which is surely the normal course of events. Of course, this leads us directly back to the preface paradox.

3.2 Dealing with the Preface Paradox

There are various desiderata that an adequate solution to the preface paradox should be capable of meeting. On the one hand, we want to continue to take measures to avoid errors from propagating by treating inconsistency as a defect for a set of commitments, which makes the afflicted assertions incapable of functioning as an uncontroversial starting point for further inquiry. On the other, Foley (1993: 117) seems to be right that it is a desideratum for any decent theory that agents should not be deemed irrational for

recognizing their own fallibility. Indeed, it seems that, if anything, it is part of being an epistemically responsible agent to do just that. Furthermore, we want to avoid the absurd outcome that the set of commitments undertaken in a book by epistemically responsible agents ends up not being suitable as a starting point for further inquiry by our standards due to its inconsistency.

In meeting these constraints we will use reflections about what the function is of the various parts of a book as our clue. As it turns out, the resulting approach ends up fitting nicely with Spohn's observation that the problem generated by the preface paradox arises due to a mixture of epistemic perspectives.[103]

If we use the present book as our example, chapters I-V serve the function of advancing substantial claims about a number of subject matters ranging from methodological issues, the semantics of conditionals, performance on psychological experiments, and the rationality assumptions embodied in ranking theory. In contrast, the preface served the opportunity to make a statement about *the epistemological status* of the claims advanced in chapters I-V (in addition to its more rudimentary functions of advertising what is to come and acknowledging the influence of others). There is thus a sense in which all the substantial claims made in this book are contained within chapters I-V and that nothing of consequence about its subject matter is stated in the preface. Accordingly, if the reader wants to look up what its author thinks about some topic to challenge it, then he or she should turn to chapters I-V and can safely ignore the preface. Hence, chapters I-V contain all the claims that I undertake an obligation to defend in writing this book *qua* author.

In contrast, in commenting in the preface on the epistemological status of the claims advanced in chapters I-V, I am already beginning to comment on what in the book can be used as a starting point for further inquiry. However, that is the task of the scorekeeper. So in a preface of this type, the author is already beginning to act as his own scorekeeper as it were, and it is here the source of the problems is to be located.

To disentangle the roles of these different epistemic perspectives, it is useful to take a look at what Brandom (1994: ch. 8) has to say in general about the interaction between the doxastic perspectives of the speaker and the scorekeeper. In Olsen (forthcoming), I have

laid out these matters more carefully, but for present purposes let the following brief sketch suffice. According to Brandom, it is a structural feature of the scorekeeping perspective that a principled distinction is drawn between *what is actually correct* and *what is merely taken to be correct*. He holds that this normative distinction is expressed through the use of *de dicto* and *de re* ascriptions, when attributing doxastic commitments to the speaker. That is, in describing the claims that the speaker has undertaken an obligation to defend on the basis of his assertions, the scorekeeper can either express the assertions in the speaker's own vocabulary in a form that he would acknowledge having undertaken, or he can specify which entities the speaker is *talking about* and what claims he is making *of* these entities using his own vocabulary. Of the two, the latter is the form used for making truth assessments as the following example illustrates:

> Bruja: "Pachamama will yield a poor harvest unless she is treated properly".
>
> Scorekeeper: "The Bruja is claiming *of* the earth *that* it will yield a poor harvest unless it is treated properly".

Once stated in its *de re* form, the Bruja can be treated as having made an acceptable assertion that any farmer will give his assent to; in its former *de dicto* version the scorekeeper might have had some reservations.

In making the distinction between what appears to be correct according to the doxastic perspective under assessment (i.e. the claim about Pachamama), and what is correct once this claim has received a *de re* specification, the scorekeeper needs a supply of propositions stating how the world actually is. To him, it will appear that his own collateral commitments make up this set (because why else accept these propositions unless they appeared to express how the world actually is to him). So in effect he is comparing the commitments of the doxastic perspective under assessment with his own doxastic commitments—in spite of the fact that it appears to him *as if* he is comparing what the Bruja takes to be correct to how things actually are.

Now the point of introducing this bit of Brandom's account is that it puts into a new light what the author is doing in the preface when starting to act as his own scorekeeper.

When acting as a scorekeeper in relation to foreign doxastic perspectives, the scorekeeper is bound to make some attributions of mistakes simply due to the differences in their collateral commitments. So here the scorekeeper has no problem with complying with the maxim that no agent is infallible as he will attribute mistakes to the commitments under assessment some of the time. However, when he is acting as a scorekeeper on a book written by himself, his comparisons of what the author takes to be correct with what is actually correct all end up falling out favorably as he is in effect comparing a set of propositions with itself. So in this case the maxim that no agent is infallible is violated and he cannot express a recognition of the fallibility of the author without producing the inconsistency expressed by the preface paradox. Actually, the problem is twofold. On the one hand, there is the problem of denying the proposition that every claim in the book is correct *qua* scorekeeper while simultaneously being committed to defending that very claim *qua* author. On the other, there is the problem that supposing that there is an error in the book, in spite of the fact that each claim was assessed as correct, ends up indicating that the set of propositions is error-prone that he presupposes expresses how the world actually is in his own truth assessments.

This is how things look from the author's side of the story. When we turn to his readers, the present suggestion is that they should construct two scores of commitments that they attribute to the author. The first is the author's deontic score *qua* author and it contains the propositions that the author has undertaken an obligation to defend during his treatment of the subject matter he is dealing with. In our example, this would be the propositions expressed in chapters I-V. The second is the author's deontic score *qua* acting as a scorekeeper on his own work and it contains the propositions that the author has undertaken an obligation to defend through his remarks in the preface. Of the two, the latter is guaranteed to be an inconsistent set so the afflicted propositions cannot be attributed entitlement, and the former is only inconsistent, if the author happens to produce an inconsistency in his treatment of the issues dealt with in chapters I-V.

For the author each claim in chapters I-V appears to be justified and correct and he states so in the preface. But the reader is well-advised not to be predisposed to accept all of the author's claims about the epistemological status of the claims made in the book due to

the inconsistency. Rather, the reader should weigh the author's fallibility higher than the fact that each claim in chapters I-V appears to be justified to the author. For what the author's fallibility means is exactly this: part of the time he makes claims that appear to him to be justified despite the fact that they are actually mistaken. In contrast, the author is unable to weigh the information about the epistemological status of his assertions in the same manner, if it would mean that he should stop acting on what he perceives to be a good justification for making a particular claim. What he can do is to improve his skills at evaluating and obtaining evidence, but no matter how good he gets, there will always be a point, where he just has to rely on what he perceives to be a good justification despite his continued fallibility.

So the way the present account seeks to avoid the absurd consequence that we can no longer use the claims advanced in books as our starting points for further inquiry is by demarcating the inconsistency produced by the preface to the deontic score of the author *qua* acting as a scorekeeper on his own claims. That the epistemically responsible author refuses in the preface to undertake an obligation to defend the claim that every claim in the book can be used as an uncontroversial starting point for further inquiry does not make the negation of the conjunction of all the claims in chapters I-V part of the actual claims advanced in the book. Surely, the point of writing the book was not just to present the reader with another large conjunction of claims that he should not accept.

No, the author's score *qua* author begins and ends with chapters I-V. And in relation to this set of commitments business is as usual. That is, if the author is reliable then the fact that a claim seems to him to be justified should be allowed to carry some weight. But ultimately the readers should make their own assessments of whether entitlement can be attributed to each individual claim and be prepared to make some attributions of mistakes on the basis of the author's general fallibility.

Since the attribution of inconsistency is only used as a way of stopping error from propagating, the present account moreover allows the scorekeeper to treat the author as continuing to be rational. The inconsistency in the author's score *qua* acting as a scorekeeper on his own work is only produced, because the agent is acting on incompatible obligations each of which seems eminently rational in its own right. On the one hand, he

continues to be the author of the book and is therefore committed to defend the claims advanced in chapters I-V. On the other, he is trying to give his readers some instruction in the preface about how to assess his own claims based on how he would have assessed them, if they were written by someone else. In this, the author tries to express a recognition of his own fallibility, which is surely the only responsible thing to do. Unfortunately, in attempting to combine both concerns he ends up producing an inconsistency in the second deontic score. But the fault lies with his incompatible obligations and not in his lack of rationality.[104]

As we have seen, this account is thus able to meet all of the desiderata for dealing with the preface paradox identified above.

4. The Bridge Principles and Problems 1-4

To return to MacFarlane's (manuscript) bridge principles, I extend this list by the following candidates, which are inspired by Brandom's account. As said, these bridge principles differ from those MacFarlane considers in dealing with commitments instead of belief, introducing the focus on acknowledged and consequential commitments, and in emphasizing the doxastic perspective of the speaker and the scorekeeper:

(I) If A, B, ⊨ C, then the speaker ought to see to it that if he/she acknowledges a commitment to A and a commitment to B, he/she acknowledges a commitment to C.

Commentary: the speakers' means for acknowledging a commitment to C consists in accepting challenges to A and B based on challenges to C.

(II) If A, B, ⊨ C, then if the speaker acknowledges a commitment to A and B, the scorekeeper is permitted/entitled to attribute a consequential commitment to C.

Moreover, since all relations of commitment preservation are entitlement preserving,[105] it holds that:

(III) If A, B, ⊨ C, then if the speaker acknowledges a commitment to A and B, and the scorekeeper both attributes an entitlement to A and B and a consequential commitment to C, the scorekeeper ought to attribute an entitlement to C.

(IV) If A, B, ⊨ C, then if the speaker is entitled to adopt a commitment to A and B, the speaker is entitled to adopt a commitment to C.

It is to be noticed that the deontic operator is given a wide scope over the whole conditional in (I). As a result, (I) describes the conditional task responsibility of the speaker to acknowledge a commitment to C, *if* he/she acknowledges a commitment to A and B. However, this is an obligation that can be fulfilled by either acknowledging a commitment to C or by withdrawing the commitment from A or B. So the first of Harman's problems is avoided. We can also set aside problem 3 as it has already received an extensive treatment.

(It should, moreover, be noted that principle (III) and (IV) were mainly stated for the purpose of completeness; they will play no further role in our treatment of problems 2 and 4 below.)

4.1 Dealing with Problems 2 and 4

One of the ramifications of making it the task of the scorekeeper to construct a (public) belief set for the speaker on the basis of his assertions is that problem 2 and 4 need to be addressed both from the perspective of the speaker and from that of the scorekeeper.

If we start out with the speaker's perspective, the first observation to be made is that the speaker has only adopted the conditional task responsibility to defend his commitments whenever appropriately challenged. Hence, the speaker need not worry about the excessive demand of having to defend all the consequences of his claims in the absence of scorekeepers, who are capable of identifying the corresponding consequential commitments and posing suitable challenges. However, as the knowledge of the implications grows, the speaker continues to run the risk of having to retract his earlier claims, if he cannot provide an adequate response to the novel challenges.

So to see how the speaker can fulfill the requirements of bridge principle (I) in light of problem 4, it suffices to notice that the context in which the speaker would have to

acknowledge a commitment to the logical consequence of his acknowledged commitments is, when challenges are posed to the consequential commitments as a way of challenging his acknowledged commitments. So what the speaker would need to do to comply with this bridge principle is merely to accept such challenges and be prepared to withdraw his commitment to A or B in the case the challenges to C turn out to be too severe.

Moreover, problem 2 is easily avoided. To the extent that challenges are hardly going to be based on trivial (and irrelevant) logical consequences of the speaker's acknowledged commitments, the speaker does not stand in danger of having to devote precious, cognitive resources to dealing with irrelevancies.

When we turn to the scorekeeping perspective, one way of dealing with this same problem of clutter avoidance would be to hold that "the algorithm" for adding logical consequences to the speaker's score as consequential commitments terminates, whenever its operation does not immediately contribute to the task of finding out whether entitlement can safely be attributed. That is, there will be no need for the scorekeeper to go through infinite sequences of conjuncts and disjuncts, if it is already clear from the outset that they are irrelevant for determining whether entitlement can be attributed.

This way of addressing problem 2 moreover opens up for a way to avoid being committed to revisionism about logic due to the restriction of *ex falso quodlibet* noted in section 2.1. Accordingly, one way of getting around this problem would be to hold that "the algorithm" for adding logical consequences to the score terminates for a particular set of assertions as soon as an inconsistency has been detected. For then the task of assessing whether entitlement can be attributed has already been solved, and the scorekeeper can proceed to challenge the speaker and criticize others that adopt commitments to the claims in question through deference to the speaker.

If we apply bridge principle (II) to problem 4 for the scorekeeping perspective, we notice that the task of assessing whether entitlement can be attributed does not impose excessive demands on the scorekeeper, because although the scorekeeper is *permitted* to add all the logical consequences as consequential commitments to the speaker's score—and to challenge him on this basis—he is *not required* to do so. Similarly, although the scorekeeper is *permitted* to run complete consistency checks on the speaker's score using all the logical

consequences as consequential commitments, he is *not required* to do so. Nor is he required to check every logical equivalent formulations of the speaker's acknowledged commitments.

As we have seen, the scorekeeper is entitled to take these measures to prevent error from propagating, when the speaker puts forward his assertion as something that others can use as an uncontroversial starting point for further inquiry. But the scorekeeper can, of course, refrain from fully exercising this right by not investigating *all* the logical consequences of the speaker's assertions, if he is willing to run the risk of letting an error slip in. Indeed at some point he must terminate prematurely due to the undecidability of logical consequence. But even if consequence were decidable, he would still have to terminate prematurely due to: (1) the complexity involved in discovering that A, B, ⊨ C may exceed what he would be able to process in even a lifetime given the best proof systems available, (2) the fact that there are infinitely many consequences of A & B, which cannot be investigated in a finite amount of time, and (3) his limited logical competence.[106]

Potentially the algorithm for executing this task takes the form of a fast and frugal heuristics (cf. Gigerenzer, 2010), which only adds the most salient consequential commitments that would be needed for the context of conversation. For surely there is a trade-off to be made between the cost of continuing to probe the speaker's (public) belief set by adding logical consequences and the potential cost of sometimes adopting error-prone claims through testimony.

However, this does not mean that we have to give in to problem 4, because as Levi (1991: ch. 2, 1997: ch. 1) has emphasized the important question is not, whether our actual performance succeeds in implementing the requirements of the principles of rationality. But rather whether we continue to recognize that we are in need of improvement whenever they don't. That is, to the extent that we continue to refine our abilities to detect consequential commitments through, for instance, education and technological assistance (e.g. use of computers, paper and pencil, and handbooks of tables), we express our recognition that there is a regulatory ideal that we stand under an obligation to approximate.

4.2 Three Further Constraints

In addition to the cases we have already considered, MacFarlane (manuscript: pp. 11-2) uses the following constraints to adjudicate between possible bridge principles. Since his concern is with the relationship between logical consequence and rational beliefs, we will need to consider whether something equivalent holds for the case of public commitments.

The first is *the strictness test*, which holds that for the general case, the agent has not done everything that he ought to, if he only believes *p* but not its logical consequence, *q*.

Although our first bridge principle did not capture the exact wording of this constraint, a case could be made that it managed to capture the gist of it by requiring that the speaker accepts challenges based on the logical consequences of his acknowledged commitments. At this point it is unclear whether anything further is needed or whether this conditional task responsibility already succeeds in making the relation between *p* and its logical consequences sufficiently strict.

The second is whether the proposed bridge principle is capable of getting the priority right so that we can still say that:

> We seek logical knowledge so that we will know how we ought to revise our beliefs: not just how we *will* be obligated to revise them when we acquire this logical knowledge, but how we are obligated to revise them even now, in our state of ignorance. (ibid.)

This concern arises, because if we were only normatively constrained by known logical consequences, it seems that "[t]he more ignorant we are of what follows from what, the freer we are to believe whatever we please" (ibid.), which seems to get things backwards.

More specifically, the concern in our context might be that since the speaker only has to acknowledge the logical consequence of his acknowledged commitments as consequential commitments by accepting suitable challenges, the speaker gets off the hook more easily the more ignorant his scorekeepers are. In response, it can be pointed out that the speaker's responsibility to accept such challenges does not come with an expiration date.[107] So he will continue to be liable to criticism, if his assertions are shown to be logically incoherent as our knowledge about the logical consequences grows. (Or rather, the

expiration date is the point, where we can no longer consider the agent's assertions as uncontroversial starting points for further inquiry, because our knowledge has grown too much in the intermediary time. But this does not guard the original agent from revision through ignorance, because what it means is merely that the assertions will lose their epistemic significance once the ignorance is overcome, if there was anything problematic about them in the first place.)

Moreover, it will still be possible to maintain on the basis of the present approach that we seek logical knowledge so as to prevent error from propagating. Hence, there will still be a pressure towards overcoming our state of ignorance on the present proposal. Similarly it holds for the scorekeeper that—although he is only permitted and not required to add the logical consequences as consequential commitments to the speaker's score according to bridge principle (II)—he risks contributing to the propagation of error, whenever he refrains from exercising this right. So he too is under pressure to overcome a state of ignorance.

The final constraint consists in being able to maintain that an agent, who refuses to take a stand on a logical consequence (e.g. their conjunction) of his beliefs is acting in a way that he ought not to. As we have seen, bridge principle (I) postpones the need for the speaker to take a stand on the logical consequences of his acknowledged commitments until a suitable challenge emerges, and it is this feature of the present account that ensures that excessive demands are not imposed on the speaker. But on the other hand, it is not clear why the agent should be forced to take a stance on all the logical consequences of his acknowledged commitments in the absence of a well-grounded suspicion about unmet, severe challenges. It might be *prudent* for the speaker to consider some of the most obvious logical consequences of his assertions before making them to avoid having to withdraw them immediately in the face of embarrassing challenges. But it is not obvious why it would constitute a failure of his epistemic responsibility as long as he is prepared to withdraw them if severe challenges emerge. And, of course, at that point (I) no longer licenses him to refrain from taking a stance on the logical consequences of his acknowledged commitments.

According to bridge principle (II), the scorekeeper is not required to take a stance on all the logical consequences of the speaker's acknowledged commitments. And it is this feature of the present account that ensures that excessive demands are not imposed on the scorekeeper. But here too it is unclear why it should be problematic that the scorekeeper refuses to take a stance on whether a logical consequence could be added to the speaker's score as a consequential commitment, unless there was some well-grounded suspicion that the scorekeeper might thereby contribute to avoiding the propagation of error. So here too our bridge principles don't seem to collide with MacFarlane's (manuscript) criteria of adequacy.

5. Conclusions and Future Work

It then appears that the present account is capable of handling the problematic cases that Harman (1986) discusses as well as the further constraints that MacFarlane (manuscript) considers. The significance of this contribution consists in that MacFarlane presents these various desiderata as standing in a tension and thus requiring some sort of trade-off, which has been avoided on the present account.

By theorizing about public commitments instead of beliefs, we are able to treat cases of inconsistency as harmless cases of incompatible obligations that cannot all be redeemed at once. By invoking the distinction between doxastic perspectives, and making it the task of the scorekeeper to construct a deontic score for the speaker that meets the minimal requirements of belief sets to decide whether entitlement can be attributed, we are able to drive a wedge between assessments of the speaker's rationality and assessments of which information we want to use for further inquiry.

This move allows the speaker to be rationally permitted to maintain inconsistent doxastic commitments, when confronted with conflicting requirements while allowing his scorekeepers to take measures to prevent errors from propagating. Moreover, we have seen that it comes with the further nicety that we can continue to remain uncommitted about revisionism about logic while avoid letting *ex falso quodlibet* ruin the deontic score of the

speaker by adding commitment to random propositions, whenever the speaker finds himself in situations of this kind.

An area for further investigation is a general comparison between the respective advantages and disadvantages of formulating the bridge principles in terms of public commitment or rational beliefs. It is surely of central importance when dealing with this issue that while it is not completely voluntarily what we believe (in the sense that if we really believe something, we cannot just decide to stop believing in it whenever we want (ibid: 15)), our acknowledged commitments is something that we can exercise full control over. For this reason it might be more natural to think about potentially conflicting obligations in terms of public commitments than in terms of beliefs, which would thereby restrict a central tool for dealing with inconsistencies to bridge principles formulated in terms of public commitments.[108]

The upshot of this final chapter has been that one can make the normative foundation of ranking theory more palatable by viewing it as applying to public commitments attributed in argumentative contexts instead of to beliefs in individual reasoning. As such, this approach to the problem of logical omniscience opens up for a new avenue of research in psychology. The take home message has been that if we are interested in the extent to which consistency, deductive closure, and the equivalent treatment of logically equivalent propositions provide a suitable normative foundation, we should not look at whether the participants actually succeed in complying to these norms in their own individual reasoning, but rather at the extent to which they recognize being bound by them in argumentative contexts through the justificatory challenges they pose and accept.

More generally, this reorientation connects with the work of Mercier & Sperber (2011), who have recently made an influential case that the primary function for which reasoning evolved is the production and evaluation of arguments. In support of this claim they cite a range of circumstantial evidence. Probably the most convincing of which is the finding that once the Wason selection task[109] was posed in groups, where the participants could deliberate about the solution in an argumentative context, the performance went up from the usual ca. 10 %[110] to about 70 % (and even to 80 %, when they had first been

presented with the problems on an individual basis). Moreover, this drastic improvement in performance was not merely the result of there being one individual in each group, who had come up with the correct solution and shared it with the others, as the verbal transcripts clearly show how some groups were able to jointly assemble all the pieces of the puzzle (Moshman & Geil, 1998).

Of course, such findings do not conclusively settle the issue about the evolutionary function of reasoning. But they do make it interesting to follow the approach sketched in this chapter to test whether the norms are being recognized in an argumentative setting as opposed to in individual reasoning.

From a philosophical perspective, one of the interesting corollaries of this reorientation is that it opens up for the application of the axioms of belief revision theory embodied in ranking theory to yield a formal account of the score of consequential commitments that the scorekeeper keeps on the speaker. In this context, it is interesting to observe that ranking theory was in part developed to solve the problem of iterated belief change in belief revision theory (Spohn 2012: ch. 4-5) and that Schaefer's (2012) whole dissertation is devoted to the problem of how to modify Brandom's account of defeasible reasoning to allow for the recovery of entitlement that has been defeated through the addition of further commitments. So here the former has the prospect of enriching the latter approach.

Second, the conditionalization rule in ranking theory could be applied to give us a precise account of the updating of entitlement of a deontic score. These observations all open up for new avenues of promising research.

Finally, through its explication of the reason relation, ranking theory provides a natural formal framework for making the inferential relations that Brandom uses to explicate propositional content precise, as we have seen. Based on this explication, Brandom can be viewed as advancing a global ranking-theoretic (or probabilistic) reason relation semantics. Within the present dissertation, this semantic analysis has been applied to the conditional connective. However, instead of following Brandom in proposing a general alternative to truth-conditional semantics, chapter II contained the compromise that we should integrate reason relations in the sense dimension of meaning and view both

the decoding of truth conditions and argumentative structure as core components of linguistic competence.

---

[91] Acknowledgement: this chapter profited from discussions with Wolfgang Spohn, Michael De, Lars Dänzer, Eric Raidl, and the other members of a reading group on *The Laws of Belief* at the University of Konstanz. I would also like to thank the participants at Thomas Müller's colloquium, Keith Stenning, and the audience at AISB50 for discussion.

[92] Predecessors: in a way Levi (1991: ch. 2, 1997: ch. 1) was the first to emphasize that one could make progress with respect to the problem of logical omniscience by thinking of it in terms of commitments rather than in terms of belief. Subsequently, Milne (2009) has gone down a similar route. What the present treatment adds is giving it a more Brandomian spin (which was already implicit in Milne (2009)) and by formulating bridge principles that are capable of dealing with the constraints presented in MacFarlane (manuscript).

[93] Explication of the framing effect: it has been shown that different ways of presenting the same information will give rise to different emotions, which in turn affects our judgments and decision making. Accordingly, the statement 'the odds of survival one month after surgery are 90 %' will be found more reassuring than the equivalent statement 'mortality within one month of surgery is 10 %' (Kahneman, 2012: 88). As a result, participants will respond differently to these two statements in spite of their logical equivalence.

[94] On Brandom's notion of material inferences: it should be noted that material inferences are used as a generic notion for content based inferences in the writings of Brandom. To be sure, Brandom does not accept the analytic/synthetic distinction for familiar Quinean reasons, but his notion of material inferences covers both what would otherwise be thought of as following in both of these categories. In his writings one thus not only finds examples of material inferences that sound like analytical inferences, like the example in the text, but also examples like inferring that a banana is ripe from its being yellow (Brandom, 2010: 104).

[95] Qualification concerning a weaker notion: this is, of course, the strictest explication possible of the notion of commitment preservation. In principle, one could also just demand that $\tau(q|\Gamma) = c$ for some high number, c. However, the version in the text was chosen, because Brandom says that the notion of commitment preservation generalizes the category of deductive inferences.

[96] Refinement through J-conditionalization: to allow for cases where $P(\Gamma) < 1$, the third condition could be replaced by Jeffrey conditionalization as follows: $\sum_{i=1}^{n}\left[P_{initial}(q|\gamma_i) \cdot P_{new}(\gamma_i)\right] > b$, for $P_{initial}(\gamma_i) > 0$.

[97] Inferentialism as a probabilistic reason relation semantics: by exploiting the idea from Spohn (2012: ch. 6) that p is a reason *for* q whenever $\tau(q|p) > \tau(q|\neg p)$, and that p is a reason *against* q whenever $\tau(q|p) < \tau(q|\neg p)$, the weak and the strong notions of incompatibility are treated as cases of when p is an inductive or a deductive reason *against* q, and entitlement preservation and commitment preservation are treated as cases, where the set Γ counts as an inductive or a deductive reason *for* q. This explication treats inferentialism as *a rank-theoretic (or probabilistic) reason-relations semantics*, and it is in general agreement with Dorn's (2005) account of the strength of arguments. However, this explication can only be partial, because it needs to be supplemented with Brandom's pragmatic account of the conditions under which the scorekeeper should add and subtract commitments and entitlements from the speaker's score, which Kibble (2005, 2006a, 2006b) has begun to formalize.

[98] Clarification on assertion: actually on Brandom's view making an assertion is putting forward a claim as something that the hearer can *use as a premise in his/her own reasoning* and not: putting it forward as *an uncontroversial starting point for further inquiry*. The reason why the argument was nevertheless formulated in the latter way was to bracket the issue of reductios. The point is that while reductios use the speaker's assertions as premises in one's own reasoning, the premises in reductios cannot be thought of as uncontroversial starting points for further inquiry. Rather I take it that reductios can be seen as a dialectical tool that scorekeepers use to show that there is a problem with the speaker's constellation of commitments. (I thank Michael De for forcing me to clarify this point.)

[99] On the use of conditionals in the bridge principles: MacFarlane (manuscript) says that the conditional can be read as the material implication in the formulation of these principles (at least to begin with). But I am not sure whether this is a good idea in light of the paradoxes of the material implication according to which $\neg p \therefore p \supset q$ for any arbitrary q, as it could introduce bridge principles of any arbitrary degree of absurdity for when C is not a logical consequence of A and B. Alternatively a semantics for the conditional could be preferred, where the paradoxes of the material implication are avoided.

[100] Reference: http://podcasts.ox.ac.uk/people/hartry-field.

[101] Separating a weak and a strong version: notice that it would also be possible to hold the view that the deontic score built up in the course of a conversation would be ruined completely by an inconsistency. Instead, a weaker version was put forward here, according to which entitlement is only withheld with respect to the assertions producing the inconsistency (e.g. p and q, where q entails non-p) and not with

respect to the whole deontic score. Yet, repeated instances of such failures can diminish one's trust in the agent, which is why the idea of blacklisting recurrent sinners is introduced below.

[102] Explication of the liar paradox: one version of the liar paradox runs as follows. The second sentence in this endnote is not true. Suppose the second sentence is true, then it is true that the second sentence is not true, and so the second sentence must not be true. Suppose it is not true, then things are as the second sentence says they are, and so it must be true.

[103] Reference: personal communication.

[104] Parallel to Moore's paradox: in exhibiting this difficulty in asserting something about one's own doxastic perspective that one would be able to assert about a foreign doxastic perspective, the preface paradox bears some similarity to Moore's paradox, which consists in that we cannot assert sentences such as '*p*, but I do not believe that *p*' or '*p*, but I believe that non-*p*' without it sounding paradoxical—in spite of the fact that it is perfectly possible for any agent that *p* is the case and that this agent either believes that non-*p* or fails to believe that *p* (cf. Brandom 1994: 605). In both cases we seem to be faced with things that we know hold with respect to any other doxastic perspective (and *a fortiori* to our own), but that we cannot assert directly about our own doxastic perspective. Perhaps the best that the author can do is to restrict himself to counterfactuals about how he would have acted as a scorekeeper if the book had been written by somebody else.

[105] Caveat concerning the reason relation explications: the explication in section 2.1 did not quite capture this feature of Brandom's account by adding the requirement that $P(\Gamma) > a$ on entitlement preservation, which found no parallel in the explication of commitment preservation. So this is one of the senses in which it was only offered as a first approximation. Another related sense in which it is only offered as a first approximation is that it does not yet contain a formal representation of a commitment to *p*. Yet, one might argue that just as a formal representation of entitlement had to be part of the explication of entitlement preservation, so a formal representation of commitment has to be part of the explication of commitment preservation.

[106] Acknowledgement: I thank Michael De for helping me to clarify this point.

[107] On commitments without an expiration date: as the practice of defending the works of deceased philosophers shows, the deontic score of an agent can outlive his biological time in virtue of other agents stepping in and administering the commitments of a deceased agent either as he would have been disposed to or in the way that would have been most optimal.

[108] Extension: however, a comparative discussion would, *inter alia*, have to compare the present bridge principles formulated in terms of public commitments with those formulated in terms of beliefs advanced in MacFarlane (manuscript) and H. Field (2009).

[109] Explication of the Wason selection task: in this task, the participants are presented with four cards, which have D, K, 3, and 7 respectively faced up and given the conditional rule 'If there is a D on one side of any card, then there is a 3 on its other side'. The task then consists in determining which cards to turn over to decide, whether the rule is true or false. To check for its falsity, the participants would have to select the D and the 7 card. Yet, most tend to select D and 3 (Manktelow, 2012: ch. 3).

[110] Reference: Evans & Over (2004: 74).

VI

Conclusion

As this dissertation has been a piece of interdisciplinary research, it is useful to take a bird's eye perspective on the respective contributions to philosophy and psychology it has made, and once more emphasize the systematic connections that unify its various parts. In chapter I, methodological recommendations were given for how to do interdisciplinary philosophy in a way that is useful to scientific purposes. As was emphasized, the point of these recommendations was neither to limit the area of legitimate projects in philosophy nor to suggest that this is the *only* way that one can conduct interesting, interdisciplinary philosophy. Rather the recommendations are to be read as a hypothetical imperative that presents a means to the end of applying philosophical theories to the scientific discourse in a way, where experimentalists will also be able to see a utility in the often speculative, theoretical discussions in philosophy. In the course of that chapter, an argument was laid out that it doesn't merely suffice to produce a theoretically interesting new theory that is compatible with existing evidence, if it is unclear what its experimentally distinguishable predictions are. The reason given was that more uncertainty is introduced with expanding the hypothesis space. So unless the newly added hypotheses have some prospect of contributing to new discoveries of their own (e.g. their own falsification), the effect of their

addition will be to leave us further away from being able to make an empirically grounded decision among the now expanded set of serious possibilities on the basis of the existing evidence. Moreover, as unique or experimentally distinguishable predictions will not benefit us much, if they can be sacked at strategically favorable times, the requirement was introduced that these predictions should ideally be hard to vary by being systematically based on the core theses of the theories in question.

Chapters II-IV implemented these recommendations with respect to the ranking-theoretic approach to conditionals by first theoretically motivating a relevance approach to conditionals in a comparative discussion of the theories of conditionals currently under consideration in psychology of reasoning (chapter II). Second, by exploiting a parallel between two-sided ranking functions and logistic regression to formulate a mathematical model based on the ranking-theoretic approach to conditionals, which allows us to derive precise, quantitative predictions of the latter for the conditional inference task (chapter III). And finally by identifying the unique and hard to vary predictions among the predictions derived in chapter III, by discussing issues of compatibility with existing empirical data, and by showing how the model formulated could be used to extending the dual source approach in psychology of reasoning (chapter IV).

In chapter V, a synoptic view was taken on the idealizing assumptions of rationality embodied in ranking theory (and other theories from formal epistemology) in the context of the rationality debates in psychology of reasoning by putting forward one strategy for making them more palatable.

As the suppositional theory of conditionals in psychology of reasoning has been the main candidate, which the relevance approach has been held up against, it is only appropriate, if its main shortcomings are summarized. In chapter II, it was found that: (a) the suppositional theory is unable to invalidate the paradoxes of the material implication for the right reason, (b) that the intuitively appealing ideas of conditionals serving as inference tickets and conditional inferences, where one is only prepared to assert the consequent conditional on the truth of the antecedent, are not systematically related to the core claims of the suppositional theory in spite of the fact that some of its proponents use these ideas to motivate it, and that (c) the suppositional theory of conditionals is incapable

rendering failures of the epistemic relevance of the antecedent for the consequent genuine semantic defects despite the fact that this failure ropes the inflicted conditionals of the cognitive utility that they would otherwise be capable of having. In chapter III, it was moreover found to be a comparative advantage of the ranking-theoretic implementation of the relevance approach that it was capable of rendering AC and DA valid, in contrast to the suppositional theory of conditionals, as both these inferences have reasonable uses.

The substantial philosophical theses that have been advanced throughout the course of the dissertation are listed below.

In appendix 1, it was suggested that we view revision of (epistemic) modal intuitions as part of the normal course of science and the speculative suggestion was advanced that such revisions could be represented formally as cases of adding atoms to the sampling space of serious hypotheses. Taken in themselves, these suggestions are ripe with implications for Bayesian epistemology and philosophy of science as they suggest that we can escape the problem of having to include all metaphysically possible hypotheses in our hypotheses space and assign a probability from the interval (0,1) to allow for subsequent updating on the basis of evidence by means of conditionalization. Here the alternative suggestion is that we operate with a restricted sampling space of serious possibilities representing the theoreticians' modal (epistemic) intuitions for the purpose of serious inquiry, which is in itself capable of being replaced as the theoretical discourse evolves. In this way, we avoid: (1) having to assume *modal omniscience* on the behalf of the theoreticians in question (consisting in having an overview of the set of all possibilities in advance), and (2) burdening them with the challenging task of always having to interpret the evidence in light of every possibility—no matter how farfetched or irrelevant it may appear to be. As the sampling space is itself subject to revisions, we avoid the problem of being stuck with assignments of 1 or 0 probabilities, which cannot be updated on through simple conditionalization.

In chapter II, a general argument was given for why lack of relevance should be considered a genuine, semantic defect of conditionals, which should not be excluded from the semantic analysis for pragmatics to deal with. In the course of this argument, the thesis was advanced that we should include reason relations among the sense dimension of

meaning dealing with its cognitive utility. Furthermore, the idea was put forward that we should introduce *the dialectical compositionality of an argument* in addition to the truth-conditional compositionality of a sentence and that there is a class of expressions, which were designated as utterance modifiers, which contribute to the former but which are not standardly thought of as contributing to the latter.

An attempt was moreover undertaken to account for the compositionality of conditionals in a systematic way by viewing compound conditionals as introducing logical connectives that operate on propositions about reason relations. Finally, the perceived objective purport of indicative conditionals was accounted for in terms of aiming at stating truths about propositions stating reason relations, when asserting an indicative, and having factual disputes about such propositions. It was then shown what options were available for providing a justification of this perceived purport on the basis of the ranking-theoretic explication of the reason relation by regimenting such disputes in either a subjectivistic or an objectivistic manner.

In chapter IV, we encountered the interesting case (from a philosophy of science perspective) of having two mathematically equivalent models, which were argued not to be equivalent in their psychological plausibility and theoretical utility as they could be distinguished along the following dimensions: (1) the psychological plausibility of the parameters of the model, (2) the underlying semantics of the conditionals and how they would model the presence of the rule in the conditional inference task, (3) the capacity of the model to generate new, interesting research questions, (4) the capacity of the model to throw new light on existing theoretical puzzles, and (5) the ability of the model to inspire researchers to derive new, interesting predictions.

In appendix 2, alternatives to Spohn's taxonomy of reason relations were put forward to accommodate the observations that: (a) supererogatory reasons are capable of being sufficient reasons, (b) necessary reasons are capable of being insufficient reasons, and (c) deductive reasons should always be sufficient reasons (independently of whether the propositions that they are reasons for are accepted beforehand).

In chapter V, an argument was given that it would be possible to deal satisfactorily with a number of constraints on formulating bridge principles between logic and norms of

reasoning discussed in the literature, if these bridge principles are formulated in terms of public commitments to answer justificatory challenges rather than in terms of belief.

Some of the nice features of the bridge principles suggested were that: (i) they allowed us to drive a wedge between assessments of rationality and assessments of the information that we want to use as a starting point for further inquiry (which in turn allowed us to deal satisfactorily with cases such as the preface paradox, where the agent is acting rationally but inconsistently), and (ii) they allowed us to prevent the principle of *ex falso quodlibet* to ruin an agent's doxastic score without thereby incurring a commitment to revisionism about logic.

From a psychological point of view, the main achievement of the dissertation is having formulated a mathematical model for the conditional inference task, which allows us to derive quantitative predictions from the ranking-theoretic approach to conditionals. Throughout chapters III and IV, the theoretical implications of this model were drawn out through a comparative discussion with the suppositional theory of conditionals and the dual-source approach. There a number of constructive suggestions were made such as how: (i) the model may help us to understand the cognitive processes underlying performing the Ramsey test (which is currently a major unsolved problem in psychology of reasoning), (ii) the logistic regression model could be used to provide an account of the content mode of reasoning that would extent the dual source approach, (ii) the logistic regression model provides the prospect of explaining the unique pattern of endorsement rates of MP, MT, AC, and DA, where all are endorsed but to different degrees, by employment of predictor relationships that hold in both directions but to different degrees, (iv) the logistic regression model is capable of retrodicting the effect of the presence of the conditional rule found in Klauer *et al.* (2010), and how (v) the model made use of a different implementation of epistemic relevance than the delta-p rule, which has only found scant empirical support in the few experiments that have directly investigated its relationship to probability assignments to conditionals. However, it was also found that although the logistic regression model predicts the main finding in the covert and overt experimental paradigms with disablers and alternative antecedents, it produces predictions for AC and MT, which have not been supported. As intuitive sense could be made of these predictions,

and as they could also be derived from Bayes' theorem, it was suggested that we view the participant's lack of sensitivity to them as a failure of rationality that may be related to the confirmation bias.

Moreover, an appendix was added to chapter IV, which shows how to model the uptake of conditional information in a way that satisfies Douven's criteria of adequacy.

Finally, the approach put forward in chapter V to dealing with the problem of logical omniscience had the implications for psychology of reasoning that we should not test, whether ordinary participants comply to the idealizing assumptions of our formal models (i.e. deductive closure, consistency, and treating logically equivalent sentences equally) in their performance in their own individual reasoning. Rather the suggestion was made that we should test, whether they recognize being bound by these norms in the justificatory challenges that they pose and accept. This suggestion opens up for a new avenue of empirical research.

References

Adams, Ernest (1965), 'The Logic of Conditionals', in: *Inquiry*, 8 (1-4): 166-97.

Andrews, K. (2012), *Do Apes Read Minds? Toward a New Folk Psychology*. Cambridge, MA: The MIT Press.

Bach (1997), 'The Semantics -Pragmatics Distinction: What It Is and Why It Matters', in: *Linguistiche Berichte* 8: 33-50

---(2006), 'Pragmatics and the Philosophy of Language', in: Horn, L. & Ward, G. (ed.), *The Handbook of Pragmatics*, Wiley-Blackwell: ch. 21.

Backlund, L. G., Bring, J., Skånér, Y., Strender, L., & Montgomery, H. (2009), 'Improving Fast and Frugal Modeling in Relation to Regression Analysis: Test of 3 Models for Medical Decision Making', in: *Medical Decision Making*, 29: 140-8.

Beller, S. (2008), 'Deontic norms, deontic reasoning, and deontic conditionals', in: *Thinking & Reasoning*, Vol. 14 (4): 305-41.

Beller, S. and Kuhnmünch, G. (2007), ‚What causal conditioning tells us about people's understanding of causality', in: *Thinking & Reasoning*, 13 (4): 426-60.

Beller, S. and Spada, H. (2003), 'The logic of content effects in propositional reasoning: The case of *conditional reasoning with a point of view*', in: *Thinking and Reasoning*, 9 (4): 335-78.

Bennett, J. (2003), *A Philosophical Guide to Conditionals*. Oxford: Oxford University Press.

Bennett, M. R., & Hacker, P. M. S. (2003), *Philosophical Foundations of Neuroscience*. Oxford: Blackwell Publishing.

Blackburn, S. (1986), 'How can we Tell Whether a Commitment has a Truth Condition', in: Travis, C. (ed.), *Meaning and Interpretation*. Oxford: Blackwell: 201-32.

Blackmore, D. (2004), *Relevance and Linguistic Meaning: The Semantics and Pragmatics of Discourse Markers*. Cambridge: Cambridge University Press.

Brandom, R. (1994), *Making it Explicit*. Cambridge, MA.: Harvard University Press.

---(2001), *Articulating Reasons*. Cambridge, MA.: Harvard University Press.

---(2010), *Between Saying & Doing. Towards an Analytic Pragmatism*. Oxford: Oxford University Press.

Brehmer, B. (1994), 'The psychology of linear models', in: *Acta Psychologica* 87: 137-54.

Brun, G. & Rott, H. (2013), 'Interpreting Enthymematic Arguments Using Belief Revision', in: *Synthese*, 190 (18): 4041-63.

Bröder, A. (2012), 'The Quest for Take-the-Best: Insights and Outlooks From Experimental Research', in: Gigerenzer, G. & Todd, P. M., *Ecological Rationality: Intelligence in the World*. Oxford: Oxford University Press: 216-40.

Byrne, R. M. J. (1989), 'Suppressing valid inferences with conditionals', in: *Cognition*, 31: 61-83.

Carrey, S. (2011), *The Origin of Concepts*, Oxford University Press.

Carruthers, P. (draft), 'Animal minds are real, (distinctively) human minds are not', URL = http://www.philosophy.umd.edu/Faculty/ pcarruthers/

Christensen, D. (2007), *Putting Logic in Its Place: Formal Constraints on Rational Belief*, Oxford University Press.

Cummins, D. D. (1995), 'Naïve theories and causal deduction', in: *Memory and Cognition*, 23 (5): 646-58.

---(2010), 'How semantic memory processes temper causal inferences', in: Oaksford, N. & Chater, N. (ed.), *Cognition and Conditionals. Probability and Logic in Human Thinking*. Oxford: Oxford University Press: 207-17.

Dancygier, B. (1998), *Conditionals and Predictions: Time, Knowledge and Causation in Conditional Constructions*. Cambridge: Cambridge University Press.

---(2003), 'Classifying Conditionals: form and function', in: *English Language and Linguistics*, 7 (2): 309-323.

Dancygier, B. & Sweetser, E. (2005), *Mental spaces in grammar: conditional constructions*. Cambridge: Cambridge University Press.

Declerck, R. & Reed, S. (2001), *Conditionals: A Comprehensive Empirical Analysis*. Berlin: De Gruyter.

De Neys, W. (2010), 'Counterexample retrieval and inhibition during conditional reasoning: Direct evidence from memory probing', in: Oaksford, N. & Chater, N. (ed.), *Cognition and Conditionals. Probability and Logic in Human Thinking*, Oxford University Press: 197-206.

Deutsch, D. (2011), *The Beginning of Infinity. Explanations that Transform the World*. London: Penguin Books.

Dhami, M. K., & Harries C. (2001), 'Fast and frugal versus regression models of human judgement', in: *Thinking and Reasoning*, 7 (1): 5-27.

Donahoe, J. W. (2010), 'Man as Machine: A Review of *Memory and the Computational Brain: Why Cognitive Science Will Transform Neuroscience by C.R. Gallistel and A. P. King*', in: *Behavior and Philosophy*, 38: 83-101.

Dorn, G. J. W. (2005), 'Eine komparative Theorie der Stärke von Argumenten', in: *Kriterion*, 19: 34-43.

Douven, I. (2008), 'The evidential support theory of conditionals', in: *Synthese* 164: 19-44.

---(2012), 'Learning Conditional Information', in: *Mind & Language*, Vol. 27 (3): 239-63.

---(2013), 'The epistemology of conditionals', in: Gendler, T. S. & Hawthrone, J. (ed.), *Oxford Studies in Epistemology: Volume 4*, Oxford University Press: 3-33.

---(forthcoming), *The Epistemology of Indicative Conditionals*. [To be published by Cambridge University Press.]

Edgington, D. (1995), 'On Conditionals', in: *Mind*, 104: 235-327.

---(1997), 'Commentary', in: Woods, M. *Conditionals*. Oxford: Oxford University Press: 95-138.

---(2000), 'General Conditional Statements: A Response to Kölbel', in: *Mind*, Vol. 109, 433: 109-116.

---(2003), 'What if? Questions about Conditionals', in: *Mind & Language*, Vol. 18 (4): 380-401.

---(2006), 'Conditionals', in: Zalta, E. N. (ed.),*The Stanford Encyclopedia of Philosophy*. (Winter 2008 Edition).

URL = <http://plato.stanford.edu/archives/win2008/entries/conditionals/>.

Eid, M., Gollwitzer, M., & Schmitt, M. (2010), *Statistik und Forschungsmethoden*, Beltz.

Evans, J. St. B. T. (2002), 'Logic and human reasoning: An assessment of the deduction paradigm', in: *Psychological Bulletin*, 128: 978-996.

---(2007), *Hypothetical Thinking: Dual Processes in Reasoning and Judgment*. New York: Psychology Press.

---(2009), 'Does rational analysis stand up to rational analysis?', in: *Behavioral and Brain Sciences*, 32: 87-88.

---(2012), 'Questions and challenges for the new psychology of reasoning', in: *Thinking & Reasoning*, 18 (1): 5-31.

Evans, J. St. B. T., Handley, S. J., & Over, D. E. (2003), 'Conditionals and Conditional Probabilities', in: *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29: 321-55.

Evans, J. St. B. T. & Over, D. (2004), *If*. Oxford: Oxford University Press.

Fernbach, P. M. & Erb, C. D. (2013), ‚A Quantitative Causal Model Theory of Conditional Reasoning', in: *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39 (5): 1327-43.

Field, A. (2009), *Discovering Statistics Using SPSS*. London: SAGE Publications. (Second Edition)

Field, H. (2009), 'What is the normative role of logic?', in: *Proceedings of the Aristotelian Society Supplementary Volume,* 83(1): 251-68.

Frankish, K. (2005), 'Non-monotonic Inference', in: K. Brown (ed.), *The Encyclopedia of Language and Linguistics*, 2nd Edition, Elsevier Science.

Frixione, M. & Lieto, A. (forthcoming), 'Concepts, perception and the dual process theories of mind'. URL = https://www.academia.edu/3319842/-Concepts_perception_and_the_dual_process_theories_of_mind

Gallagher, S. & Zahavi, D. (2008), *The Phenomenological Mind*. New York: Routledge.

Gallistel, C. R. & King, A. P. (2010), *Memory and the Computational Brain*. *Why Cognitive Science Will Transform Neuroscience*, Blackwell Publishing.

Garcia-Retamero, R. & Dhami, M. K. (2009), 'Take-the-best in expert-novice decision strategies for residential burglary', in: Psicothema, 21: 376-81.

Gauker, C. (2005), *Conditionals in Context*. Cambridge, MA.: A Bradford Book.

Gigerenzer, G. (1988), 'Woher kommen Theorien über kognitive Prozesse?', in: *Psychologische Rundschau* 39: 91-100.

---(2010), *Rationality for Mortals: How People Cope with Uncertainty*. Oxford: Oxford University Press.

Gigerenzer, G., & Brighton, H. (2009), 'Homo Heuristicus: Why Biased Minds Make Better Inferences', in: *Topics in Cognitive Science I*: 107-43.

Gigerenzer, G. & Murray, D. J. (1987), *Cognition as Intuitive Statistics*. Lawrence Erlbaum Associates.

Gigerenzer, G., Todd, P. M., & the ABC Research Group (1999), *Simple Heuristics that Make us Smart*, Oxford University Press.

Godden, D. M. & Walton, D. (2004), 'Denying the Antecedent as a Legitimate Argumentative Strategy: A Dialectical Model', in: *Informal Logic*, 24 (3): 219-43.

Goldszmidt, M. & Pearl, J. (1996), 'Qualitative probabilities for default reasoning, belief revision, and causal modeling', in: *Artificial Intelligence*, 84: 57-112.

Goodman, N. (1991) [1947], 'The Problem of Counterfactual Conditionals', in: Jackson, F. (eds.), *Conditionals*. Oxford University Press: 1-27.

Granger, R. H., & Schlimmer, J. C. (1986), ‚The Computation of Contingency in Classical Conditioning', in: Bower, G. H. (ed.), *The Psychology of Learning and Motivation*. Orlando: Academic Press, inc.: 137-92.

Gärdenfors, P. (2005), ‚Concept learning and non-monotonic reasoning', in: Cohen, H. & Lefebvre, C. (ed.), *Handbook of Categorization in Cognitive Science*. Amsterdam: Elsevier: 824-45.

Harman, G. (1986), *Change in View: Principles of Reasoning*. Cambridge, MA.: The MIT Press.

Howell, D. C. (1997), *Statistical Methods for Psychology*, Duxbury Press. (4th Edition)

Huber, F. (2013), 'Formal Representations of Belief', in: Zalta, E. N.,*The Stanford Encyclopedia of Philosophy*. (Summer 2013 Edition) URL = <http://plato.stanford.edu/archives/sum2013/entries/formal-belief/>.

Johnson-Laird, P. N. (2008), *How we Reason*. Oxford: Oxford University Press.

Juslin, P., Nilsson, H., Winman, A., and Lindskog, M. (2011), 'Reducing cognitive biases in probabilistic reasoning by the use of logarithm formats', in: *Cognition*, 120: 248-267.

Kahneman, D. (2012), *Thinking, Fast and Slow*. London: Penguin Books.

Kaplan, D. (draft), 'What is Meaning? Explorations in the theory of *Meaning as Use*'. [Unpublished manuscript dated 1997.]

Karelaia, N. & Hogarth, R. M. (2008), 'Determinants of Linear Judgment: A Meta-Analysis of Lens model Studies', in: *Psychological Bulletin*, 134 (3): 404-26.

Kee, F. Jenkins, J., McIlwaine, S., Patterson, C., Harper, S., & Shields, M. (2003), 'Fast and Frugal Models of Clinical Judgment in Novice and Expert Physicians', in: *Medical Decision Making*, 23: 293-300.

Khlentzos, D. (2004), *Naturalistic Realism and the Antirealist Challenge*. Cambridge, MA.: The MIT Press.

Kibble, R. (2005), 'Beyond BDI? Brandomian commitments for multi-agent communication', in: *Proceedings of Normative Multi-Agent Systems Workshop*, AISB Symposium, University of Hertfordshire.

---(2006a), 'Reasoning about propositional commitments in dialogue', in: *Research on Language & Computation*. Dordrecht: Springer.

---(2006b), 'Speech acts, commitment and multi-agent communication', in: *Computational & Mathematical Organization Theory*. Dordrecht: Springer.

Klauer, K. C. (manuscript), 'An INUS theory of causal conditional reasoning'.

Klauer, K. C., Beller, S., and Hütter, M. (2010), ‚Conditional Reasoning in Context: A Dual-Source Model of Probalistic Inference', in: *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Vol. 36, No. 2: 298-323.

Kölbel, M (2000), 'Edgington on Compounds of Conditionals', in: *Mind*, Vol. 109, 433: 97-108.

Levi, I. (1991), *The Fixation of Belief and its Undoing*. Cambridge: Cambridge University Press.

---(1997), *The Covenant of Reason*. Cambridge: Cambridge University Press.

Levin, I. P., Schneider, S. L., & Gaeth, G. J. (1998), ‚All frames are not created equal: a typology and critical analysis of framing effects', in: *Organizational Behavior and Human Decision Making Processes*, 76: 149-88.

Liu, I. (2003), 'Conditional reasoning and conditionalization', in: *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29: 694-709.

Lycan, W. G. (2001), *Real Conditionals*, Oxford University Press.

MacFarlane, J. (2010), 'Pragmatism and Inferentialism', in: Weiss, B. & Wanderer, J. (ed.), *Reading Brandom. On Making it Explicit*. London: Routledge: 81-95.

---(draft), 'In What Sense (If Any) Is Logic Normative for Thought?', URL = http://johnmacfarlane.net/work.html

Machery, E. (2009), *Doing Without Concepts*. Oxford: Oxford University Press.

Manktelow, K. (2012), *Thinking and Reasoning: An Introduction to the Psychology of Reason, Judgment and Decision Making*, Psychology Press.

Mares, E. D. (2007), *Relevant Logic: A Philosophical Interpretation*. Cambridge: Cambridge University Press.

Mareschal, D. (2010), 'Concepts and Inferences in the Developing Brain'. The BJEP Current Trends Conference Series 2010: Educational Neuroscience. Birkbeck College, London. 23-24 June. Lecture.
(http://www.educationalneuroscience.org.uk/files/BJEP-CEN%20Conference/DM%20Concepts_CEN.pdf)

Markman, A. B. (1999), *Knowledge Representation*. New York: Psychology Press.

Markovits, H. & Schroyens, W. (2007), 'A Curious Belief-Bias Effect. Reasoning with False Premises and Inhibition of Real-Life Information', in: *Experimental Psychology*, 54 (1): 38-43.

Martins, A. C. R. (2005), 'Theoretical Omniscience: Old Evidence and New Theory'. [Preprint] URL: http://philsci-archive.pitt.edu/id/eprint/2458

Mercier, H. & Sperber, D. (2011), 'Why do humans reason? Arguments for an argumentative theory', in: *Behavioral and Brain Sciences*, 34: 57-111.

Merin, A. (1999), 'Information, Relevance, and Social Decisionmaking: Some Principles and Results of Decision-Theoretic Semantics', in: Moss, L. S., Ginzburg, J., & de Rijke, M. (ed.), *Logic, language and computation* Vol 2. Stanford CA: CSLI Publications: 179-221.

Milne, P. (2009), 'What is the normative Role of Logic', in: *Aristotelian Society Supplementary Volume*, 83 (1): 269-98.

Moshman, D. & Geil, M. (1998), 'Collaborative Reasoning: Evidence for Collective Rationality', in: *Thinking and Reasoning*, vol. 4 (3): 231-48.

Murphy, G. (2004), *The Big Book of Concepts*. Cambridge, MA.: A Bradford Book.

Neth, H., & Beller, S. (1999), 'How knowledge interferes with reasoning – suppression effects by content and context',, in: Hahn, M. & Stoness, S. C. (ed.), *Proceedings of the Twenty First Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum, pp. 468- 473.

Nichols, S. & Stich, S. P. (2003), *Mindreading: An Integrated Account of Pretence, Self-Awareness and Understanding Other Minds*. Oxford: Oxford University Press.

Nickerson, R. S. (1998), 'Confirmation Bias: A Ubiquitous Phenomenon in Many Guises', in: *Review of General Psychology*, Vol. 2 (2): 175-220.

Nozick, R. (2001), *Invariances: The Structure of the Objective World*. Cambridge, MA.: Belknap Press.

Oaksford, M., & Chater, N. (2003), ‚Conditional Probability and the Cognitive Science of Conditional Reasoning', in: *Mind and Language*, 18: 359-79.

---(2007), *Bayesian Rationality: The Probabilistic Approach to Human Reasoning*, Oxford University Press.

---(ed.) (2010a), *Cognition and Conditionals. Probability and Logic in Human Thinking*, Oxford University Press.

---(2010b), 'Cognition and conditionals: An introduction', in: Oaksford, N. & Chater, N. (ed.), *Cognition and Conditionals. Probability and Logic in Human Thinking*. Oxford: Oxford University Press: 3-36.

---(2010c), 'Conditional inference and constraint satisfaction: Reconciling mental models and the probabilistic approach', in: Oaksford, N. & Chater, N. (ed.), *Cognition and Conditionals. Probability and Logic in Human Thinking*. Oxford: Oxford University Press: 309-333.

Oberauer, K., Weidenfeld, A., & Fischer, K. (2007), ‚What Makes Us Believe a Conditional? The Roles of Covariation and Causality', in: *Thinking and Reasoning*, 13: 340-69.

Oberauer, K. & Wilhelm, O. (2003). The meaning(s) of conditionals -conditional probabilities, mental models, and personal utilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 29: 680-639.

Olsen, N. S. (2014), 'Philosophical Theory-Construction and the Self-Image of Philosophy', in: *Open Journal of Philosophy*, Vol. 4 (3). (Scheduled for publication in August.)

---(forthcoming), 'Brandom, TCA, and the Social Foundation of Objectivity, II'.

Over, D. E., Hadjichristidis, C., Evans, J. St. B. T., Handley, S. J., & Sloman, S. A. (2007), 'The probability of causal conditionals', in: *Cognitive Psychology*, 54: 62-97.

Pampel, F. C. (2000), *Logistic Regression. A Primer*. Thousand Oaks Sage: Publications.

Pfeifer, N. (2002). *Psychological investigations of human nonmonotonic reasoning with a focus on System P and the conjunction fallacy*. URL = http://www.pfeifer-research.de/pdf/PsyHNmrNP.pdf

Pfeifer, N., & Kleiter, G. D. (2011), ‚Uncertain deductive reasoning', in: Manktelow, Over, D. E., & Elqayam, S. (Eds.), *The Science of Reason: A Festschrift for Jonathan St B. T. Evans*, Psychology Press: 145-66.

Platt, J. R. (1964), 'Strong Inference', in: *Science*, Vol. 146 (3642): 347-53.

Politzer, G. & Bonnefon, J. F. (2006), ‚Two Varieties of Conditionals and Two Kinds of Defeators Help Reveal Two Fundamental Types of Reasoning', in: *Mind & Language*, 21 (4): 484-503.

Politzer, G., Over, D. & Baratgin, J. (2010), 'Betting on Conditionals', in: *Thinking & Reasoning*, 16 (3): 172-97.

Rescorla, R. A. (1988), 'Pavlovian conditioning: It's not what you think it is', in: *The American Psychologist*, 43 (3): 151-60.

Rey, G. (1983), 'Concepts and stereotypes', in: *Cognition*, 15: 237-62.

---(1985), 'Concepts and conceptions: A reply to Smith, Medin and Rips', in: *Cognition*, 19: 297-303.

---(2005) ‚'Philosophical Analysis as Cognitive Psychology: the Case of Empty Concepts', in: Cohen, H. & Lefebvre, C. (eds.), *Handbook of Categorization in Cognitive Science*: Elsevier: 71-89.

Riemer, N. (2010), *Introducing Semantics*. Cambridge: Cambridge University Press.

Rott, H. (1986),'Ifs, Though, and Because', in: *Erkenntnis* 25: 345-70.

Ryle, G. (1950), "If', 'So', and 'Because", in: Black, M. (ed.), *Philosophical Analysis*. Ithaca: Cornell University Press: 323-40.

Saeed, J. I. (2003), *Semantics*, Blackwell Publishing. (Second Edition)

Samuels, R., Stich, S., & Bishop, M. (2002), ‚Ending the Rationality Wars: How to Make Disputes About Human Rationality Disappear', in: Elio, R. (eds.), *Common Sense, Reasoning and Rationality*. New York: Oxford University Press: 236-68.

Schaefer, R. (2012), *Brandom's Account of Defeasible Reasoning: Problems and Solutions*. (Dissertation at the University of Guelph). URL = https://atrium.lib.uoguelph.ca/xmlui/bitstream/handle/10214/3541/Reiner%20Schaefer%20PhD%20Thesis.pdf?sequence=4.

Singmann, H, & Klauer, K. C. (2011), ‚Deductive and inductive conditional inferences: Two modes of reasoning', in: *Thinking & Reasoning*, 17 (3): 247-81.

Smith, E. E. (1989), 'Three Distinctions about Concepts and Categorization', in: *Mind & Language* Vol. 4 (1-2): 57-61.

Smith, E. E., Medin, D. L., & Rips, L. J. (1984), 'A psychological approach to concepts: Comments on Rey's "Concepts and stereotypes"', in: *Cognition* 17: 265-74.

Smith, L., & Gilhooly, K. (2006), 'Regression Versus Fast and Frugal Models of Decision-Making: The Case of Prescribing for Depression', in: *Applied Cognitive Psychology*, 20: 265-74.

Smolin, L. (2008), *The Trouble with Physics: The Rise of String Theory, The Fall of a Science and What Comes Next*. London: Penguin Books.

Sperber, D., Cara, F., & Girotto, V. (1995), 'Relevance theory explains the selection task', in: *Cognition* 57: 31-95.

Spohn, W. (2006), 'Issac Levi's Potentially Surprising Epistemological Picture', in: Olsson, E. (eds.), *Knowledge and Inquiry: Essays on the Pragmatism of Isaac Levi*. Cambridge: Cambridge University Press, pp. 125-42.

---(2009), 'A Survey on Ranking Theory', in: Franz Huber (ed.), *Degrees of Beliefs*. Springer: 185-228.

---(2012), *The Laws of Belief. Ranking Theory and its Philosophical Applications*, Oxford University Press.

---(2013a), 'A ranking-theoretic approach to conditionals', in: *Cognitive Science*, 37: 1074-1106.

---(2013b), '50 Jahre Gettier: Reichen Vielleicht', in: Ernst, G. & Marani, L. (ed.), *Das Gettierproblem. Eine Bilanz nach 50 Jahren*. Münster: Mentis Verlag, pp. 179-198.

---(forthcoming), 'Conditionals: A Unifying Ranking-Theoretic Perspective'.

Stalnaker, R. C. (1999), *Context and Content*. Oxford: Oxford University Press.

Stanford, K. (2013), 'Underdetermination of Scientific Theory', in: Zalta, E. N. (ed.), *The Stanford Encyclopedia of Philosophy*. (Winter 2013 Edition). URL = <http://plato.stanford.edu/entries/scientific-underdetermination/>.

Stanovich, K. E. (2011), *Rationality & the Reflective Mind*. Oxford: Oxford University Press.

Stein. E. (1996), *Without Good Reason. The Rationality Debate in Philosophy and Cognitive Science*. Oxford: Clarendon Press.

Strawson, P. F. (1986), 'If' and '⊃", in: Grandy, R. & Warner, R. (eds.), *Philosophical Grounds of Rationality: Intentions, Categories, Ends*. Oxford: Clarendon Press, pp. 229-42.

Sturma, D. (2006), 'Ausdruck von Freiheit. Über Neurowissenschaften und die menschliche Lebensform', in: Sturma, D. (eds.), *Philosophie und Neurowissenschaften*. Berlin: Suhrkamp Verlag, pp. 187-214.

Tedeschi, J. T, & Felson, R. B. (1994), *Violence, Aggression, and Coercive Actions*. Washington, DC: American Psychological Association.

van Deemter, K. (2012), *Not Exactly: In Praise of Vagueness*. Oxford: Oxford University Press.

van der Gaag, L. C., Renooij, S., Schijf, H. J. M., Elbers, A. R., & Loeffen W. L. (2012), 'Experiences with Eliciting Probabilities from Multiple Experts', in: Greco, S., Bouchon-Meunier, Coletti, G., Fedrizzi, M., Matarazzo, B., Yager, R. R. (editors), in: *Proceedings of the Fourteenth International Conference of Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU)*: 151-60. (Communcations in Computer and Information Sciences, Vol. 299).

Varela, F. J., Thompson, E. & Rosch, E. (1991). The Embodied Mind. Cognitive Science and Human Experience. Cambridge, MA.: The MIT Press.

Verbrugge, S., Dieussaert, K., Schaeken, H. S., William, V. B. (2007), ‚Pronounced inferences: A study of inferential conditionals', in: *Thinking & Reasoning*, 13(2): 105-33.

Williamson, T. (2008), *The Philosophy of Philosophy*. Oxford: Wiley-Blackwell.

Witteman, C., & Renooij, S. (2003), 'Evaluation of a verbal-numerical probability scale', in: *International Journal of Approximate Reasoning*, 33: 117-31.

Woods, M. (1997), *Conditionals*. Oxford: Oxford University Press.
(Edited by: Wiggins, D. Commentary by: Edgington, D.)