

Philosophische Analyse / Philosophical Analysis

Herausgegeben von/Edited by
Herbert Hochberg, Rafael Hüntelmann,
Christian Kanzian, Richard Schantz, Erwin Tegtmeier

Volume / Band 59

Morality in Times of Naturalising the Mind

Edited by
Christoph Lumer

DE GRUYTER

Contents

Introduction

- Morality in Times of Naturalising the Mind – An Overview.....3
CHRISTOPH LUMER

Part I: Free Will, Responsibility and the Naturalised Mind

- 1 Naturalizing Free Will – Empirical and Conceptual Issues45
MICHAEL PAUEN
- 2 Libet's Experiments and the Possibility of Free Conscious
Decision63
CHRISTOPH LUMER
- 3 The Effectiveness of Intentions – A Critique of Wegner105
CHRISTOPH LUMER

Part II: Naturalising Ethics? – Metaethical Perspectives

- 4 Neuroethics and the Rationalism/Sentimentalism Divide.....127
MASSIMO REICHLIN
- 5 Experimental Ethics – A Critical Analysis145
ANTONELLA CORRADINI

Part III: Naturalised Ethics? – Empirical Perspectives

- 6 Moral Soulfulness & Moral Hypocrisy – Is Scientific Study
of Moral Agency Relevant to Ethical Reflection?165
MAUREEN SIE

Part IV: Neuroethics – Which Values?

- 7 The Rationale Behind Surgery – Truth, Facts, Values.....195
ARNALDO BENINI

- Biographical Notes on the Authors203
Name Index207

ISBN 978-1-61451-799-3
e-ISBN (PDF) 978-1-61451-801-3
e-ISBN (EPUB) 978-1-61451-939-3
ISSN 2198-2066

Library of Congress Cataloging-in-Publication Data

A CIP catalog record for this book has been applied for at the Library of Congress.

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie;
detailed bibliographic data are available on the Internet at <http://dnb.dnb.de>.

© 2014 Walter de Gruyter Inc., Boston/Berlin
Printing: CPI books GmbH, Leck
© Printed on acid-free paper
Printed in Germany

www.degruyter.com



Morality in Times of Naturalising the Mind – An Overview

CHRISTOPH LUMER

The scientific and philosophical attempts of recent decades to provide a neurophysiologic or otherwise naturalising explanation of the mind had an impact on ethics too, albeit with some delay. There are experiments to find the neurophysiologic bases of moral judgement and action. Neurophysiological and psychological studies on the causes of actions have provoked debates about the existence of our freedom of will, or responsibility and of practical rationality altogether; and experimental as well as medical interventions on the brain have led to the emergence of neuroethics of such interventions. The developments driven by neurophysiology go along with a general strengthening of the efforts to study the empirical bases of morals and moral action – apart from the physiological, also their psychological, cognitive and evolutionary bases. This naturalistic wave, in turn, has provoked comments on the interpretation of particular empirical findings as well as more general debates about the role and use of empirical information in ethics. Here the spectrum of positions ranges from unconditional naturalism, which sees such empirical research as the very aim of fully developed philosophy, through various intermediate metaethical conceptions, which defend the methodological autonomy of ethics but give empirical information a more or less important role in it, to apriorism, which views (normative) ethics as a purely conceptual matter and denies any relevance of empirical research for philosophy.

The present volume (apart from Benini's chapter on neuroethics) contributes to the latter, i.e. interpretative and metaethical, debates with chapters in a reflective spirit and often with a critical evaluation of major developments of the naturalistic enterprise in ethics. Apart from presenting the chapters of this volume, this introduction will provide some background information and orientation in the form of brief overviews and metaethical assessments of the main fields of the just mentioned empirical

research on morals and their bases: 1. neurophysiology and 2. psychology of action and decision, 3. moral physiology (i.e. the science of the neurophysiological basis of moral judgements and actions), and 4. moral psychology.

1. Neurophysiology of Action and Decision

Neurophysiology of action and decision explores the physiological mechanisms behind our decisions, actions, sense of agency and behind the (auto- or hetero-) attribution of such events – in particular: time, place and the interrelation of the respective neurological processes. More specific questions regard: What is the role of cognitive, control and suppression processes in decision? Where do these processes take place, probably in the prefrontal cortex, but exactly where and when? What is the role of emotional mechanisms (in the basal ganglia etc.) for decision and action execution? What are the physiological mechanisms of reward? How and where are goals and subgoals processed, in the frontopolar cortex? Which are the unconscious determinants of apparently free actions? What are the time and role of conscious decision? Does it occur simultaneously or later than the real physiological “decision”? Is it identical to or supervenient on the real “decision”, or has it only a secondary function, whereas the real “decision” occurs earlier and unconsciously? How are the capacity to control impulsive behaviour and certain action tendencies, like sexual, aggressive or compulsive, as well as the general level of activity influenced by neuromodulators, such as serotonin and dopamine, neurotransmitters and hormones? Are there ideomotor actions, i.e. actions that are caused by merely thinking of that action, and how do ideomotor actions function – e.g. because mere representation of actions and representation for executive purposes are materialised in the same place? Is awareness of action localised in the same place during intention formation as during action execution? What are good physiological predictors of actions? What are the neural correlates of self-ascriptions of actions? What is the mechanism of intention recognition in other subjects? Is it mediated by mirror neurons? What are the physiological correlates of dysfunction in actions, and, conversely, what are the consequences of certain brain lesions for our decisions and actions? And so on.

Most of the respective findings are not directly relevant for ethics¹ – not even for ethics in a broad, Aristotelian sense, which includes the prudential aspects of our actions – but some definitely are; and sometimes it is difficult to predict which questions in the end will have ethically relevant answers. Additionally, some findings are indirectly relevant, e.g. if physiological observation helps answer ethically relevant but essentially psychological questions like whether certain decisions are taken on a more emotional or a more rational basis.² (If we know in which brain areas the respective kind of processing takes place and if we dispose of respectively fine-grained observations of this processing, e.g. by functional magnetic resonance imaging (fMRI), we may establish the answer to the empirical emotion-reason question physiologically.)

Two complexes which are definitively relevant for ethics and have attracted more than ephemeral attention among ethicists are, first, the restriction of freedom and our capacity of control by anomalous physiological factors like frontal lobe disorder, insufficient serotonin levels (due to disease, for example), psychopathy or autism, and, second, unconscious determinants of our normal decisions or directly of our normal actions. The first complex has a directly practical relevance in jurisprudence and therefore has found much attention in law but not that much in philosophy, where, however, a general and also neurophysiologically informed theory of gradually restricted responsibility should be developed (Churchland 2002: 211-219).

The second complex instead, during the last two decades has received much attention in philosophy as well as in the general public in the sequel of neurophysiologist Benjam Libet's studies of unconscious determinants of intentions and actions (scientific

¹ Therefore, the present discussion of the neurophysiology of decision and action is very selective. Some overviews of this field are: first introduction: Ward 2006: chs. 8; 13; extensive overviews and some detailed discussion: Berthoz <2003> 2006; Jeannerod 1997; Passingham & Lau 2006; Spence 2009; Vartanian & Mandel 2011.

² Another example is the physiological confirmation of the psychological differentiation between actions caused by present intentions (the paradigm of intentional action) and habitualised, automatic, i.e. unconsciously initiated and performed actions; the actions of these two classes are caused by completely different neurophysiological pathways (Neal et al. 2011: 1429).

synthesis of these studies: Libet 1985; further main elaboration: Libet 2004). Libet claims to have found that intentions are preceded (by about 500 ms) and determined by unconscious but physiologically measurable (on the vertex and temples) electrical readiness potentials in the brain, which lead to the predetermined action if no conscious veto of the subject intervenes. Even for compatibilists, who believe that causal determinacy of decisions in itself does not exclude freedom of decision, such a finding would completely undermine our traditional theory and practice of ascribing freedom of decision and responsibility. This is so because the traditional theory and practice bet on (conscious) rational deliberation and decision, which is able to find new possibilities of action, to consider and respect relevant reasons and to critically and consciously scrutinise these possibilities and their consequences. If, however, the main “decision” is already taken in the form of a readiness potential and the conscious intention only reflects this “decision”, then the action cannot be determined by such a rational deliberation and decision and, hence, cannot be free. Since this is a fundamental ethical question Libet’s releases have led to a huge debate with thousands of publications; the present volume adds two more to them (chapters 1 and 2). Whereas a considerable portion of the commentators, including several (allegedly realistic) philosophers of mind and e.g. the neurologists Gerhard Roth and Wolf Singer, simply accept Libet’s findings as one of the definite proofs that freedom of decision is an illusion, many empirical scientists, however, have harshly criticised the methods, measurements and interpretation of Libet’s studies. And philosophers of action as well as ethicists, in addition to advancing in part similar criticisms, have found conceptual faults in these studies, like confusing an urge to act with an intention to act or overlooking general distal intentions or equating freedom with indeterminacy; probably there is no ethicist engaged in this debate who accepts Libet’s conclusion (detailed discussion and references: Pauen and Lumer, chapters 1 and 2 of this volume). The upshot of this critique is that Libet has observed only urges to act (instead of intentions), which, furthermore, in part were only artificially induced by the experiments themselves and whose timing is still entirely unclear, and that the assumed determining effect of the readiness potentials for action is not more than a methodical artefact.

In light of this devastating critique, which has completely demolished Libet's attack on traditional ideas of freedom of decision and action, it is surprising how much weight it is still given. Sometimes the impression is that some of Libet's followers accept his theory because they are fascinated by a picture of consciousness in general and of intention as well as of the self in particular as something like a computer display or the measurement display of some other machine, whose indications are produced by the machine and which can tell a bit about what is going on inside but has no functional role in the machine's operation; let us call this the "*display*" theory of consciousness, intention etc.³ The opposite, *personalist* view, of course, would not deny that consciousness is only the tip of an iceberg of an immensely complex and mighty, unconscious "underwater" structure, but it would stress that this conscious tip in large part and in many important respects effectively controls the ensemble. The critique of Libet's work, of course, does not imply that the personalist view has now been proved to be true and the display theory to be false, however Libet's findings do not contribute anything to a proof to the contrary.

2. General Psychology of Action

The current neuro-hype notwithstanding, general *psychology* of action is more directly relevant for ethics than the respective *physiology* because ethics and rationality theory normally use psychological and not physiological categories. They do this because, in the end, they have to propose directly applicable rules of action or decision, which, therefore, contain conditions whose fulfilment is (mostly) epistemically (directly) accessible to the subject – like one's own beliefs, desires, emotions, in contrast to neurophysiological states. In order to be able to propose good rules, ethics and rationality theory then need empirical information about the (sufficient) conditions or consequences of such epistemically directly accessible conditions. In particular, these theories need information as to whether their proposals are realisable, whether

³ Some exponents of this view also call its main point the "zombie theory" (Koch & Crick 2001; Clark et al. 2013), where this label is not intended to mean, as usually, that we do not have consciousness but only that this consciousness does not decide, whereas the unconscious machinery decides.

they are not realised necessarily (such that there is no choice to behave differently and a respective proposal would be nonsensical) and whether they are sufficiently good in relation to other possible proposals (hence information about possible alternatives and the consequences is needed); all of this information contains psychological concepts in the antecedent or in the consequent condition (Lumer 2007a; 2007b: sects. 6-8). Accordingly, many parts of decision psychology are highly relevant for a (prudential) rational decision theory as well as for general ethics (Lumer 2007b). Even the debate between Kantians and Humeans, whether an apriori approach can make justifications of morals motivationally relevant and influential or whether reliance on the subject's desires is indispensable and will shape the content of morals, is mostly and essentially a debate about a decision psychological question. In recent decades empirical research in this field has provided a wealth of useful information (overviews: Camerer 1995; Crozier & Ranyard 1997; Hardman 2009; Koehler & Harvey 2004; Manktelow 2012; Payne et al. 1993). One result e.g. is that deciders are very flexible with respect to the decision criterion used at a time; in a certain sense they decide how to decide, thereby considering in particular the preciseness and costs (mostly time) of the decision mode and adapting it to the importance of the current decision (Payne et al. 1993); this result could even be the blueprint for a rational decision design.

However, from the studies in decision psychology, some "destructive" results have garnered the most attention from ethicists, namely findings which allegedly show that humans decide less rationally than is usually assumed. Often agents do not follow the advice of rational decision theory to maximise expected utility (many important proofs were provided by Kahneman and Tversky (conspectus: Kahneman 2011: part IV)); they miscalculate probabilities; instead of rational calculations they use rules of thumb (Gigerenzer 2010). On the one hand, this scope of philosophers' attention sometimes seems to be a masochistic delight in the destruction of a noble self-image of humankind. On the other hand, a more in-depth examination of such results could sometimes even reveal deeper forms of rationality – like e.g. second order maximisation (i.e. optimisation of the optimising process itself) or dealing with cases where statistical justifications of maximising expected utility do not hold –, which have not yet been sufficiently

captured by philosophical rationality theories and, therefore, superficially have been branded as irrational (e.g. Buchak 2013).

Beyond decision psychology, psychology of action in the last twenty years has provided a number of results which might challenge a traditional conception of action, which is taken to be the basis of our practices of prudential decision, giving reasons, civil and moral responsibility. The theory of prudential rationality and ethics have to reply at least to the following findings. (i) *Automatic actions*: Some philosophers of action have already discussed automatic actions, in particular automatic routine actions, like eating “munchies” from a bowl in front oneself or shifting a car’s gear, several decades ago (e.g. Melden 1961: 86; 97-100; 202-203; historical overview: Pollard 2010). More recently, however, psychologists of action have found still other types of automatic actions, i.e. actions which are initiated and executed without attention, e.g. mimicking one’s interlocutor or conditioned reflexes, and shown their pervasiveness (overviews: Bargh & Barndollar 1996; Bargh & Chartrand 1999). The ethical problem with automatic actions is that they are not or at least do not seem to be caused by respective intentions – which however is required for an action in the traditional sense. (ii) *Spontaneous unconscious intention and action*: Instead of being produced by an automatism, unconscious action can also be produced by a, usually rather simple, spontaneous unconscious deliberation and intention, which react creatively to the current situation and to a very limited degree consider the pros and cons of at least two options, e.g. during a conversation to sit down on a chair vs. to remain standing, or to open the window vs. not doing anything in this respect. Although there seems to be an intention in such cases, the fact that this intention is unconscious may imply that it is not subject to critical scrutiny by our reason and hence we are not responsible for the action. (iii) *(Subliminal) priming of decisions*: There is a huge mass of experiments showing that subjects who have been exposed to, i.e. primed by, certain perceptions, which (unconsciously) activate related ideas are influenced in their later decisions by these ideas, e.g. after having worked on a language test which contained several words having to do with old age (or politeness or rudeness etc.) subjects behaved accordingly, e.g. they walked more slowly (Bargh et al. 1996; Bargh & Chartrand 1999: 466). In these experiments the priming was so inconspicuous that the subjects did not even detect that they had

been exposed to some accumulation of words of a certain semantic group; and similar effects occur if the priming words are presented subliminally, i.e. so briefly that they are not consciously perceivable at all. All this means that the later conscious intention has been influenced unconsciously. If we are regularly exposed to priming effects, then isn't the lion's share of our decisions unfree? (iv) *Unknowing one's intention and action*: Psychologist Daniel Wegner has collected a long list of empirical findings where people feel that they are willing an act that they are not doing or, conversely, are not willing an act that they in fact are doing, e.g. alien hand syndrome (because of a neurological lesion one hand seems to act autonomously), table turning in spiritistic séances, or believing that one moves one's hand though only an optical illusion makes another person's – moving – hand seem to be one's own (Wegner 2002: chs. 1-2). Wegner infers from this that we have no direct knowledge about our will's causing our actions and that our respective beliefs are cognitive constructs on the basis of the empirical information at hand (ibid. 67-69). He goes on to claim and proposes a respective model that our will (or its physical basis) only provides information about but does not cause our actions, which instead are prepared and brought about by unconscious processes (Illusion-of-conscious-and-empirical-will thesis) (ibid. 68; 96; 146; 342).

All of these findings stress the role of the unconscious in the production of action, and have contributed to the view which is called "the new unconscious" (Hassin et al. 2005) – in contrast to the "old", Freudian, motivated unconscious –, i.e. a cognitive unconscious (sometimes similar to the "unseen" and complex processing of a computer) that can account for many "higher" mental processes. Although these findings have been regarded as confirmations of the display theory and though they are, of course, challenges for a traditional conception of action and responsibility, which have to be discussed carefully, in the end careful and more targeted, differentiated discussion might also merely result in some revision of the traditional picture but not lead to a plain corroboration of the display theory. *Ad i: Automatic actions*: The traditional view of actions is intentional-causalist: actions, by definition, are caused (in the right way) by respective conscious intentions (or their physiological basis). This, however, does not imply that the causally effective intention is a singular proximal intention; it may be e.g. a general or distal intention. At least part of

the automatic actions, in particular habitualised routine actions, go back to intentions formed some – or even long – time ago and thereby fulfil the definitional conditions of an action (Lumer, forthcoming). Others instead may not fulfil the conditions but, precisely for that reason, are no longer be considered actions – without any need to change the definition of ‘action’. *Ad ii: Spontaneous unconscious intention and action:* Spontaneous unconscious actions could be a limiting case of actions. On the one hand, they are caused by (something like?) a deliberation and intention; on the other, because this deliberation and intention are unconscious, they did not pass the more thorough critical check of an attentive consciousness and hence the resulting behaviour may be something we are less responsible for and possibly not an action. (If unconscious deliberation reaches a critical point, revealing a critical feature of an action this often leads to attracting conscious attention; frequently, however, conscious consideration is required in the first place for detecting problematic points, so that unconscious deliberation cannot have a sufficiently deep critical function.) *Ad iii: (Subliminal) priming of decisions:* Subliminal influences on an intention do not question the status of the intention and action as such but they will make them less rational. Already psychoanalysis has revealed (other types of) unconscious influences on our decisions. The critical moral of this insight was that a reflective person should know about and study such possible unconscious influences to raise her level of rationality. An analogous lesson should be drawn for subliminal priming as well. *Ad iv: Not knowing one’s intention and action:* The traditional, intentional-causalist conception of action does not require that agents later remember their (comprehensive) intention or have direct, firsthand knowledge of their action.⁴ Therefore, the seemingly conflicting findings, collected by Wegner, regarding the lack of such knowledge, do not contribute anything to refuting the traditional conception. Wegner’s further, much stronger Illusion-of-conscious-and-empirical-will thesis, instead, *is* in contrast to the very idea of the intentional-

⁴ Anscombe and some of her followers, however, postulated such a direct, in particular not mediated by observation, knowledge about one’s action as a definitional characteristic of action (Anscombe 1957: §§6; 8; 16; 28). But this conception is intended to be an *alternative* to the traditional, intentional-causalist theory of action. The findings collected by Wegner help to refute this opponent of the traditional theory.

causalist conception of action. However, Wegner has absolutely no proof for this thesis – apart from a reference to Libet’s findings, which have been discussed above –, and there is evidence to the contrary, which sustains the intentional-causalist view (Pauen and Lumer (chs. 1 and 3), this volume).

3. Moral Physiology

At present moral physiology is a rapidly evolving field of research. In this introduction only a once-over of it with some comments can be given to get an idea of the studies done in this field and of their ethical relevance (more detailed overviews from an ethical point of view: Levy 2009; Polonioli 2009; Reichlin, this volume: ch. 4, sect. 1).

Probably the best known results of moral physiology, which have evoked much philosophical discussion, regard the neurophysiological counterparts of moral reasoning and judgement, in particular about moral dilemmas as the Trolley Problem (Greene et al. 2001; 2004; summary: Greene 2005). While their brains were being scanned with fMRI, subjects had to decide what to do in hypothetical situations like these:

Bystander: “A runaway trolley is headed for five people who will be killed if it proceeds on its present course. The only way to save them is to hit a switch that will turn the trolley onto an alternate set of tracks where it will kill one person instead of five. Should you turn the trolley in order to save five people at the expense of one? Most people say yes.” (Greene et al. 2004: 389; nearly identical: Greene et al. 2001: 2105.)

Footbridge: “As before, a trolley threatens to kill five people. You are standing next to a large stranger on a footbridge spanning the tracks, in-between the oncoming trolley and the hapless five. This time, the only way to save them is to push this stranger off the bridge and onto the tracks below. He will die if you do this, but his body will stop the trolley from reaching the others. Should you save the five others by pushing this stranger to his death? Most people say no.” (Greene et al. 2004: 389; nearly identical: Greene et al. 2001: 2105.)

Mikhail’s figures about these scenarios are: In the Bystander dilemma 90% say they would rescue the five persons, in the Footbridge dilemma only 10% do so (Mikhail 2007: 149). This

difference is astonishing because from a consequentialist point of view both situations are *prima facie* equal: five persons are saved at the cost of one.

Greene explained the difference by an evolutionarily developed, emotionally felt inhibition to cause serious bodily harm to another subject in a personal situation (physical contact, short distance, face to face etc.), as in the Footbridge dilemma, whereas impersonally caused harm, as in the Bystander dilemma, does not elicit the inhibiting emotion. The latter type of harming was not yet possible when such emotions evolutionarily developed in our ancestors and, therefore, was not included in the naturally rejected forms of social behaviour, so that impersonal harming can be decided cognitively and rationally. (Greene et al. 2004: 389-390; Greene et al. 2001; Greene 2005: 59.) Greene's "dual-process theory" adds that beside these emotional responses there are rational deliberations, which can and sometimes do outweigh the emotions; however, rational considerations need more time than the spontaneous emotional reactions (Greene 2007). Greene supports his explanation by fMRI-data: The brain regions associated with cognitive control (anterior cingulate cortex and dorsolateral prefrontal cortex) were more active in subjects when they considered the Bystander dilemma, whereas the brain regions associated with emotion and social cognition (medial prefrontal cortex, superior temporal sulcus, posterior cingulate cortex, temporal poles, amygdala) were more active when subjects considered the Footbridge dilemma. Furthermore, the minority who in the Footbridge case decided in a "utilitarian" way (sacrificing the fat man to save five) also had the emotional activation but it was counteracted by an additional higher cognitive activation; this conflict led also to longer response times in these persons. (Greene et al. 2001: 2106-2107; 2004: 390-391.) In addition, Greene has taken the fact that in the Footbridge dilemma patients with lesions of the ventromedial prefrontal cortex (VMPFC), who lack emotional inhibition against antisocial and irrationally short-sighted behaviour endorse the "utilitarian" judgement and decide faster than normals (Ciaramelli et al. 2007; Koenigs et al. 2007) as a further confirmation of his explanation (Greene 2007).

Now some ethicists have tried to make a normative ethical point of these findings in moral physiology. Peter Singer uses these results to argue against intuitionist approaches in general and against

intuitionist objections to utilitarianism in particular. According to Greene's explanation, the majority's contrasting intuitions about cases of saving five by sacrificing one in the Bystander and in the Footbridge dilemma reduce to the difference when the respective way of killing has been invented (before or after the evolutionary introduction of emotional barriers against killing fellow men), which, of course, is morally irrelevant (Singer 2005: 247-248). The longer reflection time of the utilitarian minority in the Footbridge dilemma as well as the participation of more cognitive brain areas in their decisions show that the utilitarian decision is more rational and, hence, ethically to be preferred (ibid. 349-351). Greene associates himself with Singer's argument and adds to it that deontologism, which sustains the majority view in the Footbridge dilemma (i.e. it forbids pushing the fat man from the bridge) is actually based on intuitive, emotional decisions, which only later are rationalised (in the Freudian sense) by the deontological ethical theory; this theory is only a confabulation of reasons for an arational, emotional decision, based on historically arbitrary developments. Consequentialism, however, is not driven by emotion (or at least not by the sort of "alarm bell" emotion that drives deontologism), it is inherently cognitive and rational – it systematically considers all values and flexibly weighs them –, although it does have some affective component. (Greene 2008: 39; 41; 57; 59-65.) Given these origins, for Greene it is clear which ethical theory is preferable (ibid. 76).

This remarkable march through from moral physiology to normative ethics is too straight to remain uncriticised (Reichlin and Corradini, this volume (chs. 4 and 5); some intuitionist critique: e.g. Levy 2006; Sinnott-Armstrong 2008: ch. 2.1-2.2). Whereas the physiological part has been widely accepted, already the psychological theory contains the following problems (among others). 1. Even if one accepts the main idea of Greene's explanation it remains unclear how and when rational, cognitive considerations can trump emotional reactions. 2. The difference between personal and impersonal killing seems to be really important but not sufficient for explaining the subjects' responses. Mikhail e.g. has tested a series of further variations of the Trolley situation; one of them e.g. is "Drop Man", which is very similar to the Footbridge dilemma, however, in Drop Man the large stranger is standing on a trapdoor, which you can open by remote control, thus making him fall onto the tracks etc. as before. Although this is now an impersonal killing,

consent to the rescue measure (killing one for saving five) increases from 10% (Footbridge) to only 37% (Drop Man) (Mikhail 2007: 149), thus remaining still far below the 90% (Bystander) consent in the initial impersonal killing dilemma. The responses to Mikhail's other scenarios show that rejection switches to approval only bit by bit, depending on various conditions, which may be differently important for different people. Even if every subject had only *one* central reason for his decision this reason cannot be interpersonally identical, there must be several of them. However, it is more likely that most subjects reacted to several reasons, to which they gave interpersonally different importance. Many explanations, including Greene's, of the moral judgements in the Trolley scenarios, therefore, are false because they neither are able to capture the gradual switch of the judgements with intermediate percentages of consent between the two extreme scenarios (Bystander and Footbridge) nor do they explain the minority judgments. 3. In addition, the consequentialist judgements have not been explained either. Greene seems to suggest: there is exactly one "rational" way to decide morally, namely the utilitarian; once people are able to suppress the emotionally induced decision tendencies and start to deliberate really cognitively, then they arrive at utilitarian judgements. Given the many contrasts between normative ethicists, not only between utilitarians and deontologists, this is rather unlikely. A comprehensive moral psychology should explain how it comes to the various types of moral judgment. 4. Another problem of the experimental results and the psychological explanation is that the type of emotion(s) has not been surveyed and remains entirely unclear. In the Footbridge scenario subjects may be worried about possible penal consequences (if the trolley unpredictably comes to a stop before the Footbridge, pushing down the large stranger appears to be plain manslaughter); subjects may have bad conscience because their personal morals prohibit the action they are thinking about; they may feel an emotional inhibition to do something dreadful; they may feel pity for their imagined victim; etc. In particular the whole discussion mostly ignores the difference between *moral* emotions like guilt, indignation or gratification, which are caused by moral judgements, and e.g. prosocial emotions (emotions near to and at the basis of morals) like sympathy or respect for persons, creatures and valuable things. 5. Moral emotions, as just said, are caused by specific moral judgements;

hence they cannot explain these judgements, the explanation goes just the other way round. However, (allegedly intuitive) moral judgements and principles are ontogenetically acquired and strongly influenced by culture and, to an interpersonally quite different degree, by personal rational considerations as well as by prosocial emotions and motives (i.e. emotions and motives near to morals) (Lumer 2002: 182-186; for intercultural and socio-economic differences in moral judgements: Haidt et al. 1993). There may be also an anthropological, e.g. emotional, basis of morality; but this basis has to be identified; and certainly it does not lead directly to predefined moral criteria or even singular judgements but only via long cognitive processes, which have to be investigated in much more detail. In any case, moral emotions are no evidence for a fixed natural mechanism; they may even be the result of rational reflection about moral principles, which then have been adopted and now cause the respective emotions. Therefore, Greene has no strong argument to generally discount ethicists' (deontological or even consequentialist) reflections as (Freudian) rationalisations (cf. Greene 2008: 68-69). 6. The "utilitarian" decisions of VMPFC patients cannot simply be the consequence of a lack of (social) emotions such that rationality alone determines their judgements. In the Ultimatum Game (explanation below, in sect. 4), where, among others, indignation and personally costly punishment is tested, VMPFC patients seek (revenge driven) retaliation *more* than normals (Moll & Oliveira-Souza 2007).

The ethical part of Greene's (and Singer's) argument for utilitarianism and against deontologism has, of course, also been criticised. 1. Greene is careful enough not to simply deduce utilitarianism from empirical findings, because this would violate Hume's Law. Instead, he uses a strong normative, metaethical premise, namely that a cognitive, systematic and universal moral which considers all values and weighs them flexibly is better than a moral which is limited in these respects. The use of this normative premise, however, makes his argument weaker than it first may appear. Its probative force depends on this premise, which now has to be justified, something that Greene does not do. What is more, all empirical, physiological or psychological, findings have no probative force at all in his argument, they are completely irrelevant to the ethical argument because its other necessary premises are analytical judgements about the definitional qualities of utilitarian

morals (like summing up all individual utilities), in particular about its criteria for moral valuation and obligation (cf. Corradini, this volume (ch. 5)). Hence what initially seemed to be a justification of moral principles on the basis of empirical findings turns out to be an analytical argument, which has nothing to do with these findings. There is not even an attempt to overcome Hume's Law; and this in a certain sense is good news. 2. Utilitarianism is not the only moral system which satisfies Greene's adequacy condition; many other welfare ethics do so as well, e.g. prioritarianism, moderate welfare egalitarianism, leximin and Rawls' principles of justice. Greene does not show why exactly utilitarianism should be the right, rational ethical system. 3. Above Greene's model has already been criticised to the effect that it does not explain when and why some emotional process determines the moral judgement and action and when and why rational moral judgements gain predominance. The critique just raised adds a new aspect to this problem: It remains unclear why which "rational" moral principles are considered to be just, are personally adopted and how they can acquire motivational force. Of course, this question is also about the psychic basis of rational morals, which remain unanalysed. If utilitarianism (or some other welfare ethics) relies on emotions (like sympathy or respect for persons etc.) we have to study which emotion and how, as well as how this emotion or the moral principles justified by it "translate" into motivation. Here central parts of a moral psychology are entirely missing. 4. Greene's appraisal of the VMPFC patients' moral judgements is a bit surprising. Usually this lesion is considered to be devastating, in particular because these patients no longer feel emotional warnings regarding risky consequences and, therefore, are no longer able to control spontaneous impulses which lead to irrational or antisocial behaviour (Damasio 1994: chs. 8-9 (= pp. 165-222)). However, once VMPFC patients' moral judgements coincide with a utilitarian view, Greene considers the absence of the emotional brake and the patients' decision to be particularly rational. One problem with this view is – if we accept Greene's psychological theory for a moment – that the presence of the emotional inhibition against personal killing is assessed as a kind of harmful instinct, whereas one should perhaps, to the contrary, regret the absence of a natural inhibition against impersonal killing.

Another question about moral judgements which has been ardently discussed on the basis of physiological data is whether

moral judgements are intrinsically motivational, i.e. whether (a certain form of) ethical internalism is true. Some philosophers have argued that patients with lesions of the VMPFC (Roskies 2003: 55-58) or psychopaths (with various brain damages (Kiehl 2008)) (Deigh 1996) make more or less normal moral judgements but are not motivated to act on them, so that ethical internalism is empirically false. While philosophers have accepted the physiological part of this argument the philosophical interpretation remains controversial. Some have doubted and others reaffirmed that the patients' judgements were really moral judgements (cf. the contributions of Kennett & Fine, Roskies and Smith in: Sinnott-Armstrong 2008; Nichols 2002; Cholbi 2006). Another critical point in this debate is the interpretation and significance of "ethical internalism". First, most forms of ethical internalism can be rescued from falsification by weakening the respective hypothesis e.g. to a 99% statistical correlation. Second, taking the internalist claim to be an empirical hypothesis ('moral judgement actually leads to the respective motivation') and then attacking it, probably is a straw man fallacy; not even Kant held such an hypothesis. Some, in a broad sense, normative interpretation of internalism probably makes much more sense; Bernard Williams e.g. took the connection of moral demands to one's motives to be a condition of their authority (Williams 1979); another normative reading of internalism is to consider it as an adequacy condition for a valid justification of morals: if some "justification" of a moral system (under certain conditions, of course) does not lead to a respective motivation then it is not a good justification. In any case, the physiological information in this philosophical discussion plays only a minor role; it is sufficient to know that there are some forms of brain damage which leave apparent moral judgements intact but impair moral motivation.

Some other topics of moral physiology, apart from general contributions to brain mapping of mental activities related to morals (e.g. Moll et al. 2002a; 2002b; Heekeren et al. 2003), have been moral action and moral emotion. Moll and colleagues (2006) e.g. studied the brain activities during charitable donation with the help of fMRI and found that the mesolimbic reward system is engaged by donations in the same way as when monetary rewards are obtained; in addition, orbitofrontal areas, which also play key roles in more primitive mechanisms of social attachment and aversion, specifically mediate decisions to donate or to oppose societal causes; and more

anterior sectors of the prefrontal cortex, which are associated with control of impulsive behaviour and pursuing (long-term) goals, are distinctively recruited when altruistic choices prevail over selfish material interests, thus materialising a principled moral decision. When studying brain activities during Ultimatum Games, Sanfey and colleagues (2003) confirmed the role of emotions for costly moral punishing behaviour. Several fMRI studies corroborate the long suspected vicinity of amoral disgust and indignation: they have partially overlapping neural substrates (e.g. Moll et al. 2005). Finally, Rizzolatti's discovery of mirror neurons and the explanation of their functioning illuminates the physiological basis of sympathetic feelings and actions. *Mirror neurons* are called so because they have a double function. On the one hand, they are activated when we act or have certain feelings and express them externally (physiognomically, vocally or gesturally), on the other, the same neurons are activated when we perceive others who behave alike, i.e. when they move or express their feelings in that way. In the second, passive case, the perception of other persons' behaviour or emotional expression, via the mirror neurons causes a mostly invisible micro repetition of this behaviour or expression, which generates a memory based activation of the practical sense of the movement or of the emotionally feelings, thereby leading to an empathetic, i.e. felt understanding of the other person's intention or emotion. Often, under certain conditions, the latter form of empathy leads to compassion or sympathy, which, finally, may motivate benevolent action. (Overview: Rizzolatti & Sinigaglia <2006> 2008: in particular ch. 7.) While Rizzolatti investigates mainly the cognitive side of empathy, Tania Singer and colleagues study more the emotional and motivational consequences of (cognitive) empathy. They have found e.g. that in empathic pain (for others who receive electro shocks) the usual pain centres are activated but not those sensory fields which in normal corporal pain identify the bodily origin. Furthermore, for feeling empathic pain it is not necessary to see e.g. the other's face; if there are other evidences of pain mere imagination is sufficient to elicit empathic pain. Hence for evoking empathic emotion it can be sufficient to have *some* sort of information about the other's well-being; it is more important to capture the significance for the other person. (Singer et al. 2004.) With respect to empathy driven altruistic helping, physiological data confirmed what moral psychologists had found before (e.g. Coke et

al. 1978), namely that stronger empathic pain (as well as similar prior personal experience) increases willingness to costly helping (Hein et al. 2011).

Moral physiology has been discussed here somewhat more in detail because the general neuro-surge of the last twenty years has had its strongest impact so far within practical philosophy in metaethics, in particular in the discussion about the foundations and the justification of morals. The hitherto presented explanatory models of moral judgement or action, including the most famous, i.e. Greene's model, are much too simplistic and therefore easy to falsify; this probably will change in the future with more targeted studies and more precise methods of inquiry. But so far moral *psychology* has provided much more fine-grained explanations than moral physiology. It is really astonishing that moral physiologists mostly ignore the psychological results. Moreover, the ethical importance of the physiological findings sketched here is very limited. The apparent immediate relevance of Greene's model for the decision between deontological and consequentialist morals de facto did not obtain. Empirical information about psychopaths or persons with acquired sociopathy (VMPFC-patients) is ethically important; but the psychological information about them (about their exact mental capacities and disabilities) is ethically more relevant than the physiological explanation. Similar assessments hold for Moll & colleagues' findings about the neural bases of moral decisions and emotions (psychological decision models already told us e.g. that moral considerations make up one group of aspects in general multi-attribute decisions) or for the physiological explanation of empathy (that empathy exists and can cause sympathy and then benevolent motivation, of course, has long been investigated in psychology). There are justifiable doubts that the direct relevance of neurophysiological findings for ethics will increase with advanced research. Similar reasons for the lesser importance of neurophysiology, as they have been mentioned above in the discussion of the physiology of action, hold for moral physiology as well. The main concern of ethics, as a piece of *practical* philosophy, is to answer the question 'What shall I / we do from a moral point of view?' and thereby to influence our decisions in a free way and into a moral direction. This is possible only by submitting "material", considerations, reasons, which can affect our deliberation in a non-coercive way because they fit to the kind of mental processes and

variables present in deliberation. Now, deliberative decisions are taken via mental attitudes like desires and beliefs. Hence, to influence decisions in a free, non-manipulative way (and in a moral direction) we need to know the way of functioning or the psychology of moral decision, in particular the possibilities and limits of influencing by information, enlightenment and rational reflection – ethicists are limited to these measures, they are not neurologists who want to repair or remodel brain structures –, how and which information under which conditions changes decisions and the ways of deciding. Ethicists need this type of psychological knowledge to obtain an overview of the various ways of judging and deciding, to be able to reckon with the inalterabilities of our ways of deciding and in order to be able to develop and propose the morally best among those ways of deciding which are reachable by providing information and arguments. Hence the directly needed knowledge is psychological, it is about and (at least primarily) uses the categories of what is subjectively accessible.

However, there is a role for moral physiology in ethics as well, but it is a secondary, ancillary role. In order to go beyond the recognition of behavioural relations and to reveal the phenomenal psychic processes, psychology is dependent on introspective reports (in a very broad sense). However, these reports cannot be quantitatively precise; in addition, aimed introspection interferes with the processes to be observed. If, one day in the future, we have rather precise general mappings of mental on physiological processes and have still much more detailed physiological in-vivo observation techniques at our disposal, then physiological data may help provide much more precise psychological analyses – e.g. of how intense some feeling was and how and how strongly it influenced some decision. A second role of moral physiology is explanatory; moral physiology, one day, will explain the moral psychological laws. Of course, our mental experience is only the surface of the workings of a mighty unconscious machinery (which does not exclude that main decisions and settings of the future course take place on this level), and the leaps between successive phenomenal experiences, in the end, can be explained only physiologically. However, this kind of physiological knowledge will help to understand the mental processes e.g. during deliberation, whereas probably only psychological knowledge can be used to design morally good and cognitively accessible ways of deciding.

4. Moral Psychology

The main objects of inquiry in moral psychology are moral actions and decisions, moral motives, moral emotions and moral judgements – where, however, “*moral*” sometimes (apart from “moral judgement”) is meant in a broad sense that includes actions which conform to morality and also includes decisions and emotions which systematically lead to actions conforming to morality but which are not *guided* by moral principles. Moral motives or emotions in the narrow sense are motives and emotions respectively caused by moral judgements;⁵ moral decisions and actions in the narrow sense in turn are (mainly) caused by moral motives or emotions in the narrow sense or by moral judgements. (Humeans, of course, deny that moral judgements can, as the main cause, effect actions. But this is an empirical hypothesis not an analytical stipulation.) Moral psychologists have always hypothesised that one or the other of these phenomena is prior with respect to the others in the sense of determining the others’ content. Rationalists, for example, take moral judgements to be prior to the other phenomena; in Hume’s psychology sympathy is the leading element, in Schopenhauer’s it is compassion. Presently we are witnessing an emotivist surge, according to which moral emotions (in the broad sense) determine the content of moral judgements and motivation (see below). In order to make this hypothesis comprehensible some current studies of the single objects of moral psychology have to be considered.

Let us start with *moral decision and action*. Many ethicists presuppose that moral judgement (more or less) determines moral action, so that it would be sufficient to elicit the right moral judgement to make people act in the morally right way. Empirical evidence, however, shows that moral judgement and action are quite independent (Nunner-Winkler 1999). The reason for this is that decision psychology of moral actions does not differ from that of other actions, i.e. it is a pondering of pros and cons of various options in the mould of rational decision theory, where moral considerations make up only one of the relevant aspects and have to

⁵ For some other proposals of defining ‘moral emotion’ see e.g.: Prinz & Nichols 2010: 119-120. The definition used here is a narrower version of their second definition.

be “represented” by respective motives or desires (Lynch 1978; Heckhausen 1989: 301-302). Nor have psychologists found traces of a bipartite decision system, as hypothesised by Kant (e.g. <1785> 1977: BA 36-37 / 1903, IV: 412-413), i.e. where apart from this decision theoretic, instrumentalist decision mode there also exists a second decision mode determined by the laws of reason.⁶ Altogether, however, the psychology of moral and immoral decision is somewhat neglected in current research.

Motives for acting morally can be differentiated into several main groups. Apart from 1. motives which coincidentally conform to moral requirements (e.g. good pay for a humanitarian job), there are 2. motives of rational cooperation, i.e. desires to improve social reactions to one’s own actions (in particular avoiding punishment and receiving reward or mutual cooperation) or to obtain advantages, which can only or better be reached by cooperation, 3. self-transcendent motives to further and care for some object (person, collective, place, artefact, institution, ideal etc.) different from oneself but to which one feels attached – as in love or affection, creative expansion by means of one’s works, or collectivism and pride in one’s community and culture –, 4. (general) prosocial motives which aim at other beings’ well-being or flourishing without presupposing an already existing personal relationship (in particular sympathy or compassion and respect for persons, other living beings or things felt to be valuable in themselves), and 5. moral motives (in the narrow sense), i.e. motives which have their origin in a moral judgement (cf. Lumer 2002: 169-182). Different approaches to justifying morals have been based on different groups of these motives. Game theoretical foundations of ethics and contractualism of the Hobbesian line are based on motives of rational cooperation, an ethics of caring makes recourse to self-transcendent motives, certain forms of moral sentimentalism and Schopenhauer’s theory rest upon general prosocial motives, and moral rationalism presupposes moral motives. Correspondingly representatives of these approaches have been interested in quite different studies of motives for acting morally. Some foci of psychological research on motives for acting morally have been: Contracting for mutual advantage is the paradigm of rational cooperation. It works well when these contracts are warranted by external instances with

⁶ Discussion of several Kantian decision psychologies: Lumer 2002/2003.

sanctioning power. If, however, such an external authority is not available a sort of homo oeconomicus rationality, which seeks to cleverly maximise the satisfaction of selfish preferences, recommends cheating to get the advantages of cooperation but not to pay the price of it, which, if anticipated by both partners and under certain fairly general conditions, makes rational cooperative agreements impossible – says rational game theory. Psychological evidence, however, does not confirm this prediction. People do not behave like homini oeconomici. For one thing, their moral motives make them more honest than an homo oeconomicus; for another, retaliatory emotions make them punish cheaters, which has an additional deterrent effect; in addition, fair players to a certain degree recognise other fair players and limit cooperation to them, and because cooperators in a selectively cooperating environment are more successful than non-cooperators such cooperative behaviour has been favoured by evolution (Frank 1988; Kiesler et al. 1996; Mansbridge 1990; Parks & Vu 1994). This combination, called “strong reciprocity”, of cooperatively procuring service and punishing non-cooperation is pervasive in social life (Gintis et al. 2005). Similar results have been obtained by exploring cooperative behaviour in *Ultimatum Games*: The first player can divide a given amount of goods, usually money, between herself and a second player as she pleases. However, then comes the second player’s turn. If he accepts the division both players receive the goods as assigned by the first player; if he does not accept the distribution both get nothing. If the second player were a homo oeconomicus he would accept any distribution proposed by the first player which gives him more than zero percent because even one percent is better for him than nothing. However, this is not what has been observed. For one thing, second players usually accept only offers which at least approach the equal distribution of 50% to 50%; i.e. they really pay for punishing an unfair first player, e.g. by rejecting a 20% offer. They do this out of indignation and driven by a revenge motive. For another, first players mostly do not make very low offers in the first place, because of their fairness ideals or because they fear their proposal will be rejected. (Fehr & Gächter 2001; Henrich et al. 2004.)

So there must be non-selfish motives. Self-transcendent motives often are altruistic and mostly are important supporters of acting morally. However, since many of them – though not all, think e.g. of

a person whose life project is to care for the needy – are bound to definite individual persons, small groups or limited projects and hence are not universalistic, so that they might not define what is moral, they have not found much interest among present-day ethicists. Among prosocial motives empathy driven benevolence, unlike respect for persons and things, has instead been the object of much psychological research. In particular the question whether this kind of benevolence is really altruistic or only egoism in disguise – e.g. I help you because I want to terminate my distress from seeing you suffering – has been studied thoroughly; in a series of ingenious experiments Daniel Batson has excluded at least the most common selfish explanations for the majority of subjects (e.g. Batson & Oleson 1991; overview: Stich et al. 2010). The motivational mechanism in these cases is that empathic cognition generates sympathetic emotion, which in turn induces a (motivating) intrinsic (i.e. non-instrumental) emotion-bound desire for the other's improved well-being.⁷ Of course, this does not exclude that, additionally, one hedonistically and hence in the end selfishly tries to optimise one's sympathetic feelings (i.e. minimise pity and maximise shared joy) by helping others. After all, it feels better not to live among miserable people. Analogous double mechanisms of (i) emotions inducing new and emotion-dependent intrinsic desires besides (ii) hedonistically striving for optimising one's emotions seem to exist for many *moral motives in the narrow sense* like conscientious motives, revenge motives or indignation motives. Guilt or bad conscience, for example, first and foremost is an emotion – or better: can be identical to two different emotions, first, a deconcretised fear of punishment or of losing affection and, second, the more mature version, a decline of self-esteem after a negative moral self-evaluation).⁸ Anticipatory (i.e. before acting) negative moral self-esteem, on the one hand, can induce an intrinsic desire to be morally good, and, on the other, it can remind us of the fact that executing the considered action would lead to a still worse

⁷ On various forms and development of empathy: Hoffman 2000. For the general mechanism of emotions inducing new intrinsic desires see: Lumer 2012.

⁸ Good or quiet conscience analogously consists of, first, peace of mind and comfort in not having to fear any punishment and, second, positive self-esteem, positive moral self-evaluation, moral satisfaction with and pride in oneself.

self-esteem, which is hedonically bad; of course, one can have the latter thought even without an already reduced self-esteem. Posterior negative moral self-esteem, on the one hand, induces intrinsic desires of redemption or self-punishment, and, on the other, it can provoke various hedonistic desires and intentions, e.g. to avoid the respective type of action in the future for hedonistic reasons or to improve one's self-esteem (and get rid of present guilt feelings) by doing particularly good action; and again (apart from getting rid of present guilt) one can also form these desires and intentions independently of a present low self-esteem. (Lumer 2002: 180-181; cf. also Prinz & Nichols 2010: 137-139.)

So moral motives in the narrow sense and prosocial motives work via respective *emotions*, which induce new intrinsic motives or which are the aim of hedonic desires. Respect for persons and valuable objects has not been the object of much attention in psychological research, whereas empathic emotions or vicarious affects have been extensively studied (see e.g. Batson & Oleson 1991; Coke et al. 1978; Hoffman 2000); some results have already been reported above. What has been neglected somewhat, though, is the fact that, apart from negative, unpleasant sympathy, pity, compassion or commiseration, there is also positive, pleasant sympathy with another sentient being's positive well-being – although positive sympathy is weaker than negative. An important feature of prosocial emotions is that in the main⁹ they do not depend on moral judgements. Therefore, they may be apt for justifying moral judgements – e.g. in such a way that the degree of a certain form of universalistic sympathy or of the underlying well-being define moral value. – Moral emotions in the narrow sense can be divided into four groups: 1. self-blame emotions, e.g. guilt, low moral self-esteem, shame; 2. self-praise emotions, such as moral pride, moral satisfaction with oneself and positive moral self-esteem; 3. other-blame emotions, like indignation, outrage, loathing, disgust or contempt; 4. other-praise emotions, such as moral admiration or appreciation.¹⁰ Most of these emotions have been investigated in social psychology; here is not the place to report the respective

⁹ There is, however, a modulatory effect of moral judgements on them; moral condemnation of a suffering person e.g. can reduce or block pity.

¹⁰ Prinz (2007: 68-86) and Prinz & Nichols (2010: 122) make a similar distinction but leave out praise-emotions.

details. A general question regards the origin of singular episodes of moral emotion. An answer to this question has already been given above in the definition of ‘moral emotion in the narrow sense’, namely that they originate from a moral judgement (which may be unconscious), with the consequence that the theory of moral emotions in the narrow sense refers to a theory of moral judgements. (An alternative hypothesis to this cognitivist view assumes that moral emotions are caused directly e.g. by perception or imagination, without intermediate cognitive judgements; however, given the sophistication and cultural diversity of moral emotional reactions such a direct causation is hardly plausible. Some moral psychologists think that the CAD theory gives an answer to this objection. The CAD theory holds that there are three main areas of moral concern: 1. *community*, which regards violation of communal codes including hierarchy, 2. *autonomy*, having to do with individual rights violations, and 3. *divinity*, regarding violations of purity-sanctity; and these three areas are aligned with three corresponding emotions: *contempt*, *moral anger* and *disgust* (Rozin et al. 1999; Shweder et al. 1997). However, the objection just mentioned turns also on this explanation: Communal codes, ideas of individual rights, and ideas of divinity are so sophisticated and interculturally different that their noncognitive functioning is highly implausible. In addition, the suggestion that the emotional background and the vicinity to amoral emotions (Rozin et al. 2009) indicates a natural origin of the triggering conditions of these emotions fails likewise because of the cultural diversity of the respective norms.)

If moral emotions in the narrow sense bear on moral judgements where do *moral judgements* come from? One tradition in philosophy, represented e.g. by G.E. Moore, William Ross or in recent times by Robert Audi, Michael Huemer and in a way also by John Rawls, sees (basic) moral judgements as philosophically unexplainable intuitions; and some intuitionists, tending towards moral objectivism, consider intuitions to be something like perceptions of objective moral truths. This position, however, is (at least) psychologically unsatisfactory because even if intuitions were philosophically or cognitively impenetrable they should be explained psychologically, at least for recognising whether and how they really represent objective moral facts – as it is done in perceptual psychology and physiology with respect to the empirical reality of our perceptions. Therefore, Sinnott-Armstrong and his coauthors

(2010) try to explain what, from the subject's perspective, is an unexplainable popping-up of an intuition; they explain it as the result of an unconscious application of a moral heuristic (in Gigerenzer's sense). One group of heuristics is moral rules; and another very important heuristic is the affect heuristic: 'If thinking about an act makes you feel bad, then it is morally wrong.' (ibid. 260). Though this explanation goes a step beyond mere intuitionism it is still unsatisfactory because it leaves open where those moral rules or the moral emotion come from.

A developmental-psychological tradition of explaining moral judgements has developed following Piaget (Piaget <1932> 1965; Kohlberg 1981; 1984; Turiel 1983; 1998; Nunner-Winkler <1998> 2011; Kagan 2008). A general characteristic of these theories is that they explain the ontogenetic development of our moral judgements as a progress of several "logical" stages by evolving general higher modes of cognition, which are applied to moral questions, e.g. the passage from concrete to more abstract and general thinking, the development of the competence to understand other persons' mental states or of the competence to understand reasons behind social rules. In times of naturalising the moral mind these approaches have been criticised, first, as ignoring the intuitive, automatic formation of moral judgements in favour of assuming conscious reflection and, second, as ignoring the primary role of affects in producing moral judgements and betting on the cognitive application of moral principles instead (e.g. Haidt 2001; Hauser <2006> 2008: 21-25; 38-39; 137). These criticisms, though, are somewhat superficial and often attack a straw man. Of course, moral judgements often pop up as intuitions and are accompanied by affects, but intuitions have their origins, which, however, may be rather cognitive instead of affective, and we have seen that moral emotions refer back to moral judgements. This reply to the critique does not mean that the theories in the developmental-psychological tradition are correct. Actually, they have several defects like (often) giving insufficient weight to prosocial motives or disregarding other sources of morality like seeking advantage in cooperation, proposing unclear stage differentiations and providing very gappy explanations. However, since adult morality is also due to cognitive development – it is no accident that moral standards of people with lower as compared to higher socioeconomic status and education are, in the mean, much more conventional and rigid (Haidt et al. 1993: 619; 624) – and since

cognitivist developmental-psychological theories are much more sophisticated in integrating various sources of morality (cognitive development, prosocial and moral motives, contractarian “logic” ...) and in explaining singular steps of moral development than fashionable present-day physiological and emotivist theories of moral judgement, the potential of those theories and of the explanatory force of rational development as one source of moral development is currently grossly underrated.

Contemporary emotivist theories of moral judgement (e.g. Haidt 2001; 2012; Haidt & Bjorklund 2008; Haidt et al. 1993; Nichols 2004; Prinz 2007; in part: Greene 2005; 2007; 2008; Greene et al. 2001; 2004) assume that moral judgements always or mostly have an emotional genesis, they are arrived at as a consequence of moral emotions (in Haidt these are emotionally felt moral intuitions including moral emotions). Of course, moral emotions have a specific range of eliciting conditions but, as these theories hold, their fulfilment leads directly to the moral emotion, which then gives rise to the moral judgement. (In cognitivist theories of moral emotions, instead, the cognitive (but not necessarily conscious) moral judgement is the eliciting condition.) Since subjects have only little access to the emotion generating process the eventual justification of a moral judgement occurs later, mostly it is a rationalisation and often a mere confabulation. Haidt also allows a very limited influence of reflection on moral intuitions, and Nichols and Prinz allow for a second mechanism of reasoning generated moral judgements; but with respect to the emotional generation of moral judgements their models tend to be rather nativist. Haidt e.g. assumes six modules for the main themes of morality (2012) and endorses the CAD theory (Rozin et al. 1999; 2009), and Nichols and Prinz tend in this direction too (Prinz & Nichols 2010: 140-141).¹¹

¹¹ Hauser (<2006> 2008) has developed a non-emotivist theory of moral judgement, which could be used by emotivists as well: the universal moral grammar theory. (Another universal moral grammar theory has been provided by Mikhail (2007; 2011).) According to this theory, we possess innate and not verbally known moral principles (prohibition of killing, injuring, cheating, stealing, breaking promises, adultery and the like (Hauser <2006> 2008: 54)), i.e. the universal moral grammar, which during socialisation is automatically adapted to the domestic culture and morality in particular by permitting exceptions to originally unrestricted prohibitions. A difference with respect to the emotivist models is that, according to Hauser's theory, after analysing the situation, in particular the agent's intentions, this

These theories have been criticised, in particular by ethicists (e.g. Fine 2006; Levy 2009: 6-7; Corradini and Reichlin, this volume (chs. 4 and 5)): The models mostly do not distinguish between prosocial and moral emotions and may perhaps fit better to prosocial than to moral emotions. The nativist tendency does not capture the cultural or even individual formation of moral norms and the cultural, individual and ontogenetic diversity and specificity of their contents (different people are e.g. indignant about quite different things). The criteria for our moral judgements and many singular cases are too complex to be processed by automatic mechanisms so that the models cannot explain many moral judgements (Haidt's model e.g. does not explain the minority views in the Bystander and Footbridge scenario, Greene's model explains only the views in these two extreme scenarios but not the answers in the intermediate Trolley scenarios). The theories mention only the topics but say next to nothing about the exact contents of the (emotional) morality. Insofar as the models admit some influence of moral reflection on moral judgements these influences are not well integrated into the main model. Fast intuitions can be the result of prior learning or of unconscious inferential processes, and they can have the status of hypotheses, which then are subject to critical scrutiny whether they can be justified – as, say, a mathematician deals with some intuition about the solution of a mathematical problem. Hence the existence of such intuitions does not say anything about the truth of the generalised intuitionism. Sticking to judgements which one cannot justify may be the consequence of having acquired the respective criteria from authorities, so that one may surmise the justifiability of these criteria but not know a real justification (higher education tendentially leads to querying authority-based principles and hence to reducing acceptance of mere authority-based principles). One may even have forgotten a justification and remember only that there was one. In Haidt's tricky cases (eating one's dead pet dog, incest between siblings who use contraceptives, masturbating with a dead

system immediately provides the moral judgement and only later is an emotion added. Some problems of this theory are these. The principles are vague, not universal and only deontological in nature (moral valuation is missing). The theory does not explain how subjects can develop individual morals. It leaves out rational designing of morality and the role of prosocial motives. Finally, Hauser does not really try to prove this theory.

chicken and later eating it etc. (Haidt et al. 1993; Haidt 2001: 814-817)) there may be sensible rules in the background whose application only in this particular situation does not make sense, so that many subjects who have only a vague idea of the reasons behind these rules become unsure. As a consequence the reactions in such tricky cases do not say that much about the normal cases of moral judgement. Altogether, the models are not based on empirical studies of conscious reflection prior to moral judgement, and they are too simplistic for explaining the many sources of and influences on morality.

Resuming the just sketched discussion, one can note that all these psychological models of moral judgements are one-sided in one way or the other, always neglecting some sources and mechanisms. Hence we need a more integrative model, which may work like this. A child's original adoption of moral standards could be heteronomous via, first, seeking social gratification (avoiding punishment for immoral behaviour and pursuing reward for good behaviour) and, second, belief in authority, i.e. the belief that the standards introduced by the socialising agents will be good and important also for the child himself. A strong changer of the standards once adopted then is cognitive progress. This entails, first, gradually understanding more complex, more general and more abstract moral standards as well as their justifications (which first may be taught to the child and adolescent but also acquired autonomously by own reflection), second, seeking coherence, i.e. trying to arrive at fewer but more comprehensive standards, which capture earlier more concrete standards, as well as trying to eliminate contradictions emerging in this process and, third, as a consequence of a more critical attitude towards authority, asking for primary justifications for moral standards adopted so far and eventually discarding them if no satisfying justification is obtained. The content of cognitive progress is neutral with respect to the content of morality. Such a content may be introduced now during ontogenesis via autonomous and universal sources of morality which do not themselves depend on already adopted moral criteria: namely prosocial motives (sympathy and respect) and rational cooperation. Finally, the moral criteria adopted in this way may lead to instances of moral judgement mainly via the usual cognitive processes of judgement formation, which also permit unconscious processing followed by intuitions, emotional emphasising of important features,

influencing “correct” cognitive processing by primes etc. (Lumer 2002: 182-186).

All in all, we have seen that moral psychology in principle, because of the type of knowledge it is trying to obtain, which among others speaks of tokens that make up our deliberation, adds much more of the empirical information we need in normative and metaethics than moral physiology. Moral psychologists have already provided many interesting results (for some more specific praise: Stich et al. 2010: 202), which deserve more attention in ethics and in moral physiology. Along these lines, moral physiology has mostly an ancillary function – Cushman et al. (2010: 47) claim e.g. that mainly neuroscientific findings revealed the participation of emotions in moral reasoning, admitting, though, that there are respective psychological evidences as well. So far physiological research has fulfilled this ancillary function only to a rather limited degree, and this is also due to a pervasive disregard of moral psychological findings by moral physiologists.

5. Overview of the Present Volume

The chapters of this volume contribute to various parts of the fields of research just outlined.

The aim of *Michael Pauen’s “Naturalizing Free Will – Empirical and Conceptual Issues”* is to show that naturalising empirical research of consciousness and agency does not undermine our self-understanding as self-conscious and responsible agents but leads to an improved understanding of these qualities. Free will is his example for proving this claim. He provides a compatibilist conception of free will as self-determination by one’s own preferences, which in turn could be effectively rejected by one’s decision; and he criticises incompatibilist conceptions as having fewer advantages and leading to requirements like ultimate control, which are self-contradictory and hence unrealisable even in an indeterminate world. The second part of Pauen’s chapter defends the reality of free will against Libet’s and Wegner’s physiological and psychological objections. One main argument against Libet’s interpretation of his results is e.g. that the decisive intention in Libet’s experiments is already formed when the subjects accept the experimenter’s specific instructions, so that a study of the readiness potentials and urges to

act prior to the single actions cannot reveal anything about the freedom of the relevant intention. Pauen's critique of Wegner's theory of the "illusion of conscious will" includes reports of some experiments which show that intentions cause respective actions.

Christoph Lumer's "Libet's Experiments and the Possibility of Free Conscious Decision" and the related "The Effectiveness of Intentions – A Critique of Wegner" are detailed critiques of these two main physiological and psychological attacks on the possibility of free will. In the Libet-chapter, after showing that the truth of Libet's interpretation of his experiments would indeed imply that even a compatibilist freedom of the will as well as actions (in the action philosophical sense) would not exist, Lumer compiles a wealth of criticisms of this interpretation, which question e.g. the temporal order of the readiness potential and the urge to act, deny that Libet has ever studied intentions (because an urge to act is not an intention), and stress the decisiveness of the prior intention. The final section broadens the defence of the existence of free will against a much more general idea of Libet's mind-time theory, according to which conscious experiences are always only the end of an amplifying process; and positively it sketches a theory of the functional role of consciousness in intention formation.

Lumer's "The Effectiveness of Intentions – A Critique of Wegner" defends the intentional-causalist conception of action (in an action an intention causes the respective behaviour in the right way) against Wegner's illusion-of-conscious-will thesis. While Lumer accepts Wegner's main idea, i.e. that our posterior knowledge about our intentions and their causing our actions rests on the constructive processing of the available empirical evidence, and takes it to be compatible with intentional causalism, he criticises those less central parts of Wegner's model which indeed are incompatible with intentional causalism and shows them to be unfounded.

Massimo Reichlin's chapter "*Neuroethics and the Rationalism/Sentimentalism Divide*" criticises Haidt's and Greene's emotivist theories of moral judgement and sketches an alternative model of moral judgement, which combines sentimentalism and cognitivism. Some of Reichlin's criticisms of the emotivist theories are that these models disregard personal identity and the ontogenesis of moral judgements as well as their practical function, and they overlook the reflective part in the formation of moral judgements: moral

judgements are the result of a reflective dealing with spontaneous moral emotions. These criticisms are then turned into positive hypotheses which delineate a model of moral judgement: Emotions, in particular sympathy, are necessary conditions for authentic moral judgements but are potentially in conflict with personal interests, so that reflection has to choose between the various options; the resulting decision then is motivated by emotion but justified by reflection. Metaethically categorised, because of the universality of the respective emotions, this leads to a sentimentalistically enriched cognitivism without moral realism.

Antonella Corradini's "Experimental Ethics – A Critical Analysis", first, defends the justificatory capacity of traditional moral philosophy in general and an intuitionist approach in ethics in particular against attacks by some experimental philosophers who, with the help of moral psychology, try to show that these intuitions are not reliable. Corradini responds in three ways: with a counterattack on the conclusiveness of the experimental results (e.g. it is not necessarily the moral part of the cognitive process that is responsible for the variations of moral intuitions); by referring to more sophisticated intuitionist approaches (of John Rawls and Richard Hare), which can deal with divergent intuitions; and by a fairly general methodological objection, namely that according to Hume's Law empirical findings cannot undermine the ethical quality of moral judgements. Second, Corradini works out methodological difficulties of the neurophysiological explanations of moral judgements, e.g. whether they really explain moral beliefs and not perhaps amoral repugnance. In particular she criticises Greene's empirically based attack on deontology: From a deontological perspective, emotions are only contingent concomitants; therefore, their presence does not say anything about the validity of deontology.

Maureen Sie in her chapter "*Moral Soulfulness and Moral Hypocrisy – Is Scientific Study of Moral Agency Relevant to Ethical Reflection?*" argues for the general claim that ethics is dependent on empirical investigation of moral agency in order to be able to suggest realistic moral aims. Mainly, however, she rejects two strong, scientifically nurtured attacks on the moral nature of our apparently moral actions and argues for a revision of the traditional picture of moral agency. First, many findings of moral psychology seem to show that the moral reasons we provide for our actions are

only confabulations so that these actions are not really moral. Sie replies that even if the reasons given later were not conscious during the decision, they usually played a role in the decision; they work unconsciously like perception does in routine movements. Second, experiments by Batson and colleagues seem to show that the motive behind moral action is not the desire to be moral but to appear so. Sie replies with a critique that the examined actions were not really morally obligatory and with a reinterpretation: We learn the contents of morals by taking part in a moral practice; the desire to appear to be moral is part of our disposition to adopt the morality of our environment.

Arnaldo Benini's chapter "*The Rationale Behind Surgery – Truth, Facts, Values*" is a contribution to neuroethics, i.e. the applied ethics of interventions in the neurological realm, from a neurosurgeon's point of view. The question Benini wants to answer is which information, values and criteria should enter into or determine the decision about medical treatment. As an example he discusses several forms of brain tumor, their consequences and risks as well as the possibilities and chances for surgical removal. Several consequences have to be pondered in the respective decisions: life expectancy, risks, pain, functional impairment etc. accompanying non-intervention on the one hand, and range and probability of life prolongation, risks, impairment of mental functions, pain, fear and nuisance etc. caused by the various forms of medical treatment. Benini makes the case for full autonomy of patients with respect to the valuation and pondering of these consequences and hence letting them decide on the basis of the best information provided by qualified physicians.

REFERENCES

- Anscombe, G[ertrude] E[lizabeth] M[argaret] (1957): *Intention*. Oxford: Blackwell.
- Bargh, John A.; Kimberly Barndollar (1996): Automaticity in Action. The Unconscious as Repository of Chronic Goals and Motives. In: Peter M. Gollwitzer; John A. Bargh (eds.): *The Psychology of Action. Linking Cognition and Motivation to Behavior*. New York; London: Guilford Press. 457-481.
- Bargh, John A.; Tanya L. Chartrand (1999): The Unbearable Automaticity of Being. In: *American Psychologist* 54. 462-479.

- Bargh, John A.; M. Chen; L. Burrows (1996): Automaticity of social behavior. Direct effects of trait construct and stereotype activation on action. In: *Journal of Personality and Social Psychology* 71. 230-244.
- Batson, C. Daniel; K. C. Oleson (1991): Current status of the empathy-altruism hypothesis. In: *Prosocial Behavior* 12. 62-85.
- Berthoz, Alain (<2003> 2006): *Emotion and Reason. The Cognitive Science of Decision Making*. Translated by Giselle Weiss. Oxford [etc.]: Oxford U.P.
- Buchak, Lara (2013): *Risk and Rationality*. Oxford: Oxford U.P.
- Camerer, Colin [F.] (1995): Individual Decision Making. In: John H. Kagel; Alvin E. Roth (eds.): *The Handbook of Experimental Economics*. Princeton, NJ: Princeton U.P. 587-703.
- Cholbi, M. (2006): Belief Attribution and the Falsification of Motive Internalism. In: *Philosophical Psychology* 19. 607-616.
- Churchland, Patricia Smith (2002): *Brain-Wise. Studies in Neurophilosophy*. Cambridge, MA; London: MIT Press.
- Ciaramelli, Elisa; Michela Muccioli; Elisabetta Ladavas; Giuseppe di Pellegrino (2007): Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex. In: *SCAN* 2. 84-92.
- Clark, Andy; Julian Kiverstein; Tillmann Vierkant (eds.) (2013): *Decomposing the Will*. Oxford: Oxford U.P.
- Coke, Jay S.; C. Daniel Batson; Katharine McDavis (1978): Empathic Mediation of Helping. A Two-Stage Model. In: *Journal of Personality and Social Psychology* 36. 752-766.
- Crozier, Ray [W.]; Rob Ranyard (1997): Cognitive process models and explanations of decision making. In: Rob Ranyard; W. Ray Crozier; Ola Svenson (eds.): *Decision Making. Cognitive Models and Explanations*. Oxford: Routledge. 3-20.
- Cushman, Fiery; Liane Young; Joshua D. Greene (2010): Multi-System Moral Psychology. In: John M. Doris; The Moral Psychology Research Group: *The Moral Psychology Handbook*. Oxford: Oxford U.P. 47-71.
- Damasio, Antonio R. (1994): *Descartes' Error. Emotion, Reason, and the Human Brain*. New York: G. P. Putnam's Sons.
- Deigh, John (1996): Empathy and Universalizability. In: Larry May; Marilyn Friedman; Andy Clark (eds.): *Mind and Morals. Essays on Cognitive Science and Ethics*. Cambridge, MA; London: MIT Pr.; Bradford Book. 199-219. – Reprinted in: John Deigh: *The Sources of Moral Agency. Essays in Moral Psychology and Freudian Theory*. Cambridge: Cambridge U.P. 1996. 160-180.
- Fehr, Ernst; Simon Gachter (2001): Fairness and Retaliation. In: L. Gerard-Varet; Serge-Christophe Kolm; Jean Mercier Ythier (eds.): *The Economics of Reciprocity, Giving and Altruism*. Basingstoke: Macmillan. 153-173.
- Fine, Cordelia (2006): Is the emotional dog wagging its rational tail, or chasing it? Reason in moral judgment. In: *Philosophical Explorations* 9. 83-98.

- Frank, Robert H. (1988): *Passions within Reason. The Strategic Role of the Emotions*. New York; London: Norton.
- Gigerenzer, Gerd (2010): *Rationality for Mortals. How People Cope with Uncertainty*. 2nd ed. New York; Oxford: Oxford U.P.
- Gintis, Herbert; Samuel Bowles; Robert [T.] Boyd; Ernst Fehr (eds.) (2005): *Moral Sentiments and Material Interests. The Foundations of Cooperation in Economic Life*. Cambridge, MA; London: MIT Press.
- Greene, Joshua D. (2005): Emotion and cognition in moral judgement. Evidence from neuroimaging. In: Jean-Pierre Changeux; Antonio R. Damasio; Wolf Singer; Y. Christen (eds.): *Neurobiology of Human Values*. Berlin; Heidelberg: Springer. 57-66.
- Greene, Joshua D. (2007): Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains. In: *Trends in Cognitive Sciences (TICS)* 11, No.8. 322-323.
- Greene, Joshua D. (2008): The Secret Joke of Kant's Soul. In: Walter Sinnott-Armstrong (ed.): *Moral Psychology. Vol. 3: The Neuroscience of Morality. Emotion, Brain Disorders, and Development*. Cambridge, MA; London: MIT Press. 35-79.
- Greene, Joshua D.; Leigh E. Nystrom; Andrew D. Engell; John M. Darley; Jonathan D. Cohen (2004): The Neural Bases of Cognitive Conflict and Control in Moral Judgment. In: *Neuron* 14. 389-400.
- Greene, Joshua D.; R. Brian Sommerville; Leigh E. Nystrom; John M. Darley; Jonathan D. Cohen (2001): An fMRI Investigation of Emotional Engagement in Moral Judgment. In: *Science* 293, 14 September 2001. 2105-2108.
- Haidt, Jonathan (2001): The Emotional Dog and Its Rational Tail. A Social Intuitionist Approach to Moral Judgment. In: *Psychological Review* 108. 814-834.
- Haidt, Jonathan (2012): *The Righteous Mind. Why Good People are Divided by Politics and Religion*. London: Allen Lane (Penguin).
- Haidt, Jonathan; F. Bjorklund (2008): Social intuitionists answer six questions about moral psychology. In: Walter Sinnott-Armstrong (ed.): *Moral Psychology. Vol. 2: The Cognitive Science of Morality. Intuition and Diversity*. Cambridge, MA; London: MIT Press. 181-217.
- Haidt, Jonathan; Silvia Helena Koller; Maria G. Dias (1993): Affect, Culture, and Morality, or Is It Wrong to Eat Your Dog? In: *Journal of Personality and Social Psychology* 65. 613-628.
- Hardman, David (2009): *Judgment and Decision Making. Psychological Perspectives*. Malden, MA: Wiley-Blackwell.
- Hassin, Ran R.; James S. Uleman; John A. Bargh (eds.) (2005): *The New Unconscious*. New York [etc.]: Oxford U.P.
- Hauser, Marc D. (<2006> 2008): *Moral Minds. How Nature Designed Our Universal Sense of Right and Wrong*. 3rd ed. London: Abacus.

- Heckhausen, Heinz (1989). *Motivation und Handeln*. 2nd, completely revised ed. Berlin [etc.]: Springer.
- Heekeren, Hauke R.; Isabell Wartenburger; Helge Schmidt; Hans-Peter Schwintowski; Arno Villringer (2003): An fMRI study of simple ethical decisionmaking. In: *Cognitive Neuroscience and Neuropsychology* 14, No 9, 1 July 2003. 1215-1219.
- Hein, Grit; Claus Lamm; Christian Brodbeck; Tania Singer (2011): Skin conductance response to the pain of others predicts later costly helping. In: *PLoS One* 6,8. E22759.
- Henrich, Joseph; Robert Boyd; Samuel Bowles; Colin Camerer; Ernst Fehr; Herbert Gintis (eds.) (2004): *Foundations of Human Sociality. Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies*. New York [etc.]: Oxford U.P.
- Hoffman, Martin L. (2000): *Empathy and Moral Development. Implications for Caring and Justice*. Cambridge: Cambridge U.P.
- Jeannerod, Marc (1997): *The Cognitive Neuroscience of Action*. Oxford: Blackwell.
- Kagan, Jerome (2008): *Morality and Its Development*. In: Walter Sinnott-Armstrong (ed.): *Moral Psychology. Volume 3: The Neuroscience of Morality. Emotion, Brain Disorders, and Development*. Cambridge, MA; London: MIT Press. 297-312.
- Kahneman, Daniel (2011): *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.
- Kant, Immanuel (<1785> 1977/1903): *Grundlegung zur Metaphysik der Sitten*. (¹1785; ²1786.) (Groundwork for the Metaphysics of Morals.) In: Idem: *Werkausgabe*. Ed. by Wilhelm Weischedel. Vol. VII. Frankfurt, am Main: Suhrkamp ²1977. 7-102. – Or in: Idem: *Kants Werke. Akademie-Textausgabe*. Vol. 4. Berlin: de Gruyter 1903. 385-464.
- Kiehl, Kent A. (2008): *Without Morals. The Cognitive Neuroscience of Criminal Psychopaths*. In: Walter Sinnott-Armstrong (ed.): *Moral Psychology. Volume 3: The Neuroscience of Morality. Emotion, Brain Disorders, and Development*. Cambridge, MA; London: MIT Press. 119-149.
- Kiesler, S.; L. Sproull; K. Waters (1996): A Prisoner's Dilemma Experiment on Cooperation with People and Human-Like Computers. In: *Journal of Social Psychology* 70. 47-65.
- Koch, Christof; Francis Crick (2001): The zombie within. In: *Nature* 411,6840. 893.
- Koehler, Derek J.; Nigel Harvey (2004) (eds.): *Blackwell Handbook of Judgment and Decision Making*. Oxford: Blackwell.
- Koenigs, Michael; Liane Young; Ralph Adolphs; Daniel Tranel; Fiery Cushman; Marc Hauser; Antonio Damasio (2007): Damage to the prefrontal cortex increases utilitarian moral judgements. In: *Nature, Letters* 446. 908-911.

- Kohlberg, Lawrence (1981): *Essays on Moral Development. Vol. I: The Philosophy of Moral Development.* San Francisco, CA: Harper & Row.
- Kohlberg, Lawrence (1984): *Essays on Moral Development. Vol. II: The Psychology of Moral Development.* San Francisco, CA: Harper & Row.
- Levy, Neil (2006): Cognitive Scientific Challenges to Morality. In: *Philosophical Psychology* 19. 567-587.
- Levy, Neil (2009): Empirically Informed Moral Theory. A Sketch of the Landscape. In: *Ethical Theory and Moral Practice* 12. 3-8.
- Libet, Benjamin (1985): Unconscious cerebral initiative and the role of conscious will in voluntary action. In: *Behavioral and Brain Science* 8. 529-566.
- Libet, Benjamin (2004): *Mind Time. The Temporal Factor in Consciousness.* Cambridge, MA; London: Harvard U.P.
- Lumer, Christoph (2002): Motive zu moralischem Handeln. In: *Analyse & Kritik* 24. 163-188.
- Lumer, Christoph (2002/2003): Kantischer Externalismus und Motive zu moralischem Handeln. In: *Conceptus* 35. 263-286.
- Lumer, Christoph (2007a): The Action-Theoretic Basis of Practical Philosophy. In: Christoph Lumer; Sandro Nannini (eds.): *Intentionality, Deliberation and Autonomy. The Action-Theoretic Basis of Practical Philosophy.* Aldershot: Ashgate. 1-13.
- Lumer, Christoph (2007b): An Empirical Theory of Practical Reasons and its Use for Practical Philosophy. In: Christoph Lumer; Sandro Nannini (eds.): *Intentionality, Deliberation and Autonomy. The Action-Theoretic Basis of Practical Philosophy.* Aldershot: Ashgate. 157-186.
- Lumer, Christoph (2012): Emotional Decisions. The Induction-of-Intrinsic-Desires Theory. In: Alessandro Innocenti; Angela Sirigu (eds.): *Neuroscience and the Economics of Decision Making.* Abingdon, UK; New York: Routledge. 109-124.
- Lumer, Christoph (forthcoming): Reasons and Conscious Control in Automatic Actions.
- Lynch, John G. jr.; Jerry L. Cohen (1978): The Use of Subjective Expected Utility Theory as an Aid to Understanding Variables That Influence Helping Behavior. In: *Journal of Personality and Social Psychology* 36. 1138-1151.
- Manktelow, Ken (2012): *Thinking and Reasoning. An Introduction to the Psychology of Reason, Judgment and Decision Making.* Hove, East Sussex; New York: Psychology Press.
- Mansbridge, Jane J. (ed.) (1990): *Beyond Self-Interest.* Chicago; London: University of Chicago Press.
- Melden, Abraham I. (1961): *Free Action.* London; New York: Routledge and Kegan Paul; Humanities Press.
- Mele, Alfred R. (2009): *Effective Intentions. The Power of Conscious Will.* Oxford: Oxford U.P., USA.

- Mikhail, John (2007): Universal Moral Grammar. Theory, Evidence and the Future. In: Trends in Cognitive Sciences 11. 143-152.
- Mikhail, John (2011): Elements of Moral Cognition. Rawls' Linguistic Analogy and the Cognitive Science of Moral and Legal Judgment. Cambridge: Cambridge U.P.
- Moll, Jorge; Frank Krueger; Roland Zahn; Matteo Pardini; Ricardo de Oliveira-Souza; Jordan Grafman (2006): Human fronto-mesolimbic networks guide decisions about charitable donation. In: Proceedings of the National Academy of Sciences of the United States of America (PNAS) 103, no. 42, Oct 9 (2006). 15623-15628.
- Moll, Jorge; Ricardo de Oliveira-Souza (2007): Moral judgments, emotions and the utilitarian brain. In: Trends in Cognitive Sciences 11. 319-321.
- Moll, Jorge; Ricardo de Oliveira-Souza; Ivanei E. Bramati; Jordan Grafman (2002a): Functional Networks in Emotional Moral and Nonmoral Social Judgments. In: NeuroImage 16. 696-703.
- Moll, Jorge; Ricardo de Oliveira-Souza; Paul J. Eslinger; Ivanei E. Bramati; Janaína Mourao-Miranda; Pedro Angelo Andreiuolo; Luiz Pessoa (2002b): The Neural Correlates of Moral Sensitivity. A Functional Magnetic Resonance Imaging Investigation of Basic and Moral Emotions. In: The Journal of Neuroscience 22, 7, 1st April 2002. 2730-2736.
- Moll, Jorge; Ricardo de Oliveira-Souza; Fernanda Tovar Moll; Fatima Azevedo Ignacio; Ivanei E. Bramati; Egas M. Caparelli-Daquer; Paul J. Eslinger (2005): The Moral Affiliations of Disgust. A Functional MRI Study. In: Cognitive and Behavioral Neurology 18,1, March. 68-78.
- Neal, David T.; Wendy Wood; Mengju Wu; David Kurlander (2011): The pull of the past. When do habits persist despite conflict with motives? In: Personality and Social Psychology Bulletin 37. 1428-1437.
- Nichols, Shaun (2002): How Psychopaths Threaten Moral Rationalism. Or, Is it Irrational to Be Amoral? In: The Monist 85. 285-304.
- Nichols, Shaun (2004): Sentimental Rules. On the Natural Foundations of Moral Judgment. Oxford [etc.]: Oxford U.P.
- Nunner-Winkler, Gertrud (<1998> 2011): The development of moral understanding and moral motivation. In: Franz E. Weinert; Wolfgang Schneider (eds.): Individual development from 3 to 12. Findings from the Munich longitudinal study. 2nd ed. New York: Cambridge U.P. 253-290.
- Nunner-Winkler, Gertrud (1999): Moralische Motivation und moralische Identität. Zur Kluft zwischen Urteil und Handeln. In: Detlef Garz; Fritz Oser; Wolfgang Althof (eds.): Moralisches Urteil und Handeln. Unter Mitarbeit von Friedhelm Ackermann. Frankfurt, Main: Suhrkamp. 314-339.
- Parks, C. D.; A. D. Vu (1994): Social Dilemma Behavior of Individuals from Highly Individualist and Collectivist Cultures. In: Journal of Conflict Resolution 38. 708-718.

- Passingham, Richard E.; Hakwan C. Lau (2006): Free Choice and the Human Brain. In: Susan Pockett; William P. Banks; Shaun Gallagher (eds.): *Does Consciousness Cause Behavior?* Cambridge, MA: MIT Press. 53-72.
- Payne, John W.; James R. Bettman; Eric J. Johnson (1993): *The adaptive decision maker.* Cambridge: Cambridge U.P.
- Piaget, Jean (<1932> 1965): *The moral judgement of the child. (Le jugement moral chez l'enfant. 1932.)* Transl. by M. Gabain. New York: Free Press.
- Pollard, Bill (2010): *Habitual Actions.* In: Timothy O'Connor; Constantine Sandis (eds.): *A Companion to the Philosophy of Action.* Chichester: Wiley-Blackwell. 74-81.
- Polonioli, Andrea (2009): *Recent Trends in Neuroethics. A Selected Bibliography.* In: *Etica & Politica / Ethics & Politics.* 9,2. 68-87.
- Prinz, Jesse J. (2007): *The Emotional Construction of Morals.* Oxford: Oxford U.P.
- Prinz, Jesse J.; Shaun Nichols (2010): *Moral Emotions.* In: John M. Doris; The Moral Psychology Research Group: *The Moral Psychology Handbook.* Oxford: Oxford U.P. 111-146.
- Rizzolatti, Giacomo; Corrado Sinigaglia (<2006> 2008): *Mirrors in the Brain. How Our Minds Share Actions and Emotions. (So quel che fai. 2006.)* New York: Oxford U.P.
- Roskies, Adina (2003): *Are ethical judgments intrinsically motivational? Lessons from acquired 'sociopathy'.* In: *Philosophical Psychology* 16. 51-66.
- Rozin, Paul; Jonathan Haidt; Katrina Fincher (2009): *From oral to moral. Is moral disgust an elaboration of a food rejection system?* In: *Science* 323. 1179-1180.
- Rozin, Paul; Laura Lowery; Sumio Imada; Jonathan Haidt (1999): *The CAD Triad Hypothesis. A Mapping Between Three Moral Emotions (Contempt, Anger, Disgust) and Three Moral Codes (Community, Autonomy, Divinity).* In: *Journal of Personality and Social Psychology* 76. 574-586.
- Sanfey, Alan G.; James K. Rilling; Jessica A. Aronson; Leigh E. Nystrom; Jonathan D. Cohen (2003): *The Neural Basis of Economic Decision-Making in the Ultimatum Game.* In: *Science* 300. 1755-1758.
- Shweder, Richard A.; Nancy C. Much; Manamohan Mahapatra; Lawrence Park (1997): *The "Big Three" of Morality (Autonomy, Community, Divinity) and The "Big Three" Explanations of Suffering.* In: Allan Brandt; Paul Rozin (eds.): *Morality and Health.* New York: Routledge. 119-169.
- Singer, Peter (2005): *Ethics and Intuitions.* In: *Journal of Ethics* 9. 331-352.
- Singer, Tania; Ben Seymour; John O'Doherty; H. Kaube; Raymond J. Dolan; Chris D. Frith (2004): *Empathy for pain involves the affective but not sensory component of pain.* In: *Science* 303. 1157-1162.
- Sinnott-Armstrong, Walter (2008) (ed.): *Moral Psychology. Volume 3: The Neuroscience of Morality. Emotion, Brain Disorders, and Development.* Cambridge, MA; London: MIT Press.

- Sinnott-Armstrong, Walter; Liane Young; Fiery Cushman (2010): Moral Intuitions. In: John M. Doris; The Moral Psychology Research Group: The Moral Psychology Handbook. Oxford: Oxford U.P. 246-272.
- Spence, Sean A. (2009): *The Actor's Brain. Exploring the cognitive neuroscience of free will.* Oxford: Oxford U.P.
- Stich, Stephen; John M. Doris; Erica Roedder (2010): Altruism. In: John M. Doris; The Moral Psychology Research Group: The Moral Psychology Handbook. Oxford: Oxford U.P. 147-205.
- Turiel, Elliot (1983): *The development of social knowledge, morality and convention.* Cambridge: Cambridge U.P.
- Turiel, Elliot (1998): The development of morality. In: W. Damon (Series ed.); N. Eisenberg (Vol. ed.): *Handbook of child psychology. Vol. 3: Social, emotional, and personality development.* 5th ed. New York: Wiley. 863-932.
- Vartanian, Oshin; David R. Mandel (2011) (eds.): *Neuroscience of Decision Making.* New York; Hove: Psychology Press.
- Ward, Jamie (2006): *The Student's Guide to Cognitive Neuroscience.* Hove; New York: Psychology Press.
- Wegner, Daniel M. (2002): *The Illusion of Conscious Will.* Cambridge, MA; London: MIT Press.
- Williams, Bernard: *Internal and External Reasons* (1979). In: Ross Harrsion (ed.): *Rational Action. Studies in philosophy and social science.* Cambridge; London: Cambridge U.P. 17-28.