

PHILOSOPHICAL IMPLICATIONS AND MULTIDISCIPLINARY CHALLENGES OF MORAL PHYSIOLOGY

Stephan Schleim and Felix Schirmann

University of Groningen

Abstract. Neuroethics deals with the normative implications of advances and new technology of neuroscience. Some scholars argue that experiments on moral judgment might allow solutions to moral problems in the future or already nowadays. We discuss this research under the label of moral physiology to delineate this theoretical question from the normative implications of applied neurotechnology. After summarizing influential theories of the field we turn to a methodological and theoretical reflection concerning the way to investigate moral judgment experimentally, as well as functional magnetic resonance imaging, one of the leading methods of behavioural and cognitive neuroscience. We relate this to general challenges within neuroethics, philosophy, and a multi-disciplinary view on human morality. We argue that moral physiology may indeed yield normatively relevant findings but only under the assumption of certain normative stances which cannot be justified ultimately by neuroscience experiments.

Keywords: neuroethics, moral cognition, moral psychology, neurophilosophy, fMRI, moral decision-making, moral emotion, moral naturalism

DOI: 10.3176/tr.2011.2.02

1. Introduction

When Roger Sperry received the Nobel Prize for Physiology or Medicine in 1981 for his work on split-brain patients and the functional specialization of the cerebral hemispheres, he concluded his Nobel Lecture with the announcement that scientific progress will soon have a far reaching impact on the values and beliefs by which humans are living (Sperry 1981a). In his view, science had an “unmatched potential for the shaping of ethical values” and “[i]n the worldview perspectives and truths of science we will find the best key to valid moral guidelines” (Sperry 1981b:3). Among the multitude of scientific disciplines, Sperry particularly had his own field in mind: brain research. Philosophies, value-systems, and religious doctrines, he explained, “will stand or fall depending on the

kinds of answers that brain research eventually reveals. It all comes together in the brain” (ibid. 4).

When reading some contemporary neuroscientific and philosophical papers on moral cognition and behaviour, one might get the impression that Sperry’s idea has prevailed thirty to forty years after his announcement. While Michael Gazzaniga, who formerly worked together with Sperry on the split-brain patients, still hopes that we might identify and live more fully by “a universal set of ethics, built into our brains” (Gazzaniga 2005:xix), William Casebeer already draws the tentative conclusion that “the moral psychology required by virtue theory is the most neurobiologically plausible” (Casebeer 2003:841) and suggests jointly with Patricia Churchland that the investigation of brain processing related to moral decisions “may allow us to eliminate certain moral theories as being psychologically and neurobiologically unrealistic” (Casebeer and Churchland 2003:171).

In a similar vein, but with a different outcome, both Joshua Greene and Peter Singer interpret neuroscientific research on decisions to sacrifice few in order to save many (Greene et al. 2001, 2004) in a way that denounces intuitions *against* utilitarianism as irrational and ultimately defend the utilitarian outcome as the rational solution on these grounds (Greene 2007, Green et al. 2004, Singer 2005). These interpretations suggest that long-debated issues in moral philosophy related to ethically right human conduct can nowadays be informed or perhaps even solved by means of brain research. Indeed, in a news feature accompanying the first original publication of this field in *Science* (Greene et al. 2001), it is implied that this research might now fulfil a function that traditionally provided “job security for philosophers” (Helmuth 2001:1971).

This new kind of research and particularly its philosophical interpretations seem to constitute a new chapter in the debate on moral naturalism, that is, the identification of moral properties with some kind of natural properties, in this case properties of brain activations. At the same time, it is a central and perhaps even *the* central issue concerning the ethical implications of neuroscience: If progress in neuroscience did not only offer new possibilities of human treatment and enhancement calling for the discussion of its ethical, legal and social aspects (Giordano and Gordijn 2010, Farah 2010, Illes 2006, Levy 2007, Nagel 2010, Racine 2010) but also direct insight into moral right and wrong, then a cultural revolution of the kind envisaged by Sperry might indeed be imminent. Both of these aspects, the ethical description and analysis of neuroscience applications and the neuroscientific investigation of moral decisions, summarized in simpler terms as ethics of neuroscience and neuroscience of ethics, have previously been subsumed under the concept of ‘neuroethics’ (Roskies 2002). Because we want to avoid confusion between the applied and theoretical questions concerning neuroscience and ethics, we use the concept of ‘moral physiology’ in the remainder of this paper. Just as ‘moral psychology’ refers to the psychological investigation of moral phenomena, ‘moral physiology’ refers to their physiological investigation with a particular focus on brain research.

Our aim in this paper is to analyze the philosophical implications and multidisciplinary challenges of moral physiology. We start out with a summary of central findings and theories in section 2. Since these investigations are based on complex conceptual, methodological, and experimental presumptions, we find it necessary to continue with a theoretical reflection in section 3 which is necessary to gain a better understanding of the experimental possibilities and implications of contemporary moral physiology. With the joint knowledge of the empirical and theoretical aspects described in sections 2 and 3 it is then possible to analyze in more detail what the philosophical implications and multidisciplinary challenges actually consist of in section 4. We finally draw a conclusion for current and future empirical, as well as theoretical research on moral physiology in section 5.

2. Contemporary moral physiology in a nutshell

Reflections on the ethics of human life and the human moral faculties have a long cultural tradition. In the younger history of the sciences and especially psychology, morality also played an important role, particularly since the early 20th century (Nadelhoffer, Nahmias, and Nichols 2010). Pertinent examples are studies within psychodynamic theory (Freud 1930), cognitive-developmental approaches (Kohlberg and Puka 1994, Piaget 1932), from the perspective of psychological situationism (Carpendale and Krebs 1992), and with an emphasis on empathy (Batson et al. 1981).

Since modern methods of brain research, particularly functional magnetic resonance imaging (fMRI), allow the investigation of brain responses associated with basic and more complex cognitive processes and are aided by powerful computational visualization techniques, virtually every aspect of the human mind that can somehow be investigated in a laboratory setting has come under neuroscientific scrutiny. Moral perception and cognition are no exceptions. The number of publications within moral physiology is steadily increasing since 2001 and has already reached such a level that it is beyond the scope of this paper to address them all individually. We thus focus on three influential theories in this section and leave aside further trials to relate moral judgments to more general capacities of social cognition (Young et al. 2007) or other kinds of normative cognition, such as legal decisions (Buckholtz et al. 2008, Schleim et al. 2011).

2.1. The dual-process theory

In the original studies of Greene and colleagues (Greene et al. 2001, 2004), participants were confronted with moral dilemmas adapted from the ethical scholarly literature (e.g. Thomson 1985, Unger 1996), such as the following: A runaway trolley approaches five workmen standing on the tracks who will certainly be killed if nothing happens. However, there is the possibility to throw a switch in order to divert the trolley onto a sidetrack, where one workman is standing who would then be killed. Dilemmas of this kind were called ‘moral impersonal’. In another

case, the situation is somewhat different: A big stranger is imagined to stand on a footbridge that is spanning the track. Instead of the possibility to divert the trolley onto a sidetrack, the option is then to push the stranger off the bridge in order to stop the train. Dilemmas of this kind were called ‘moral personal’.

The participants were asked to judge the appropriateness of a reaction to the dilemmatic situation in a forced-choice situation while their brain activation was recorded with fMRI. Responses in the ‘moral personal’ cases were associated with brain activation in ‘emotional’ brain areas such as the medial prefrontal and the posterior cingulate cortex, while the ‘moral impersonal’ responses were related to brain areas associated with working memory in the prefrontal and parietal lobes (Greene et al 2001). Accompanying the brain activations, they found some evidence suggesting an ‘emotional interference effect’, since decisions endorsing the action to sacrifice few in order to save many in the ‘moral personal’ condition took two seconds longer, on average. The central additional finding of their later study was that these decisions, finally deemed ‘utilitarian’ judgments (see section 3.2), in a subset of difficult ‘moral personal’ dilemmas were associated with higher activations in the dorsolateral prefrontal cortex (DLPFC) previously related to cognitive control (Greene et al. 2004).

Greene and colleagues integrated their findings into a dual-process view of moral judgment, claiming that two different cognitive functions, emotion and cognitive control, can be equally active when subjects are confronted with moral dilemmas. According to their model, emotional reactions elicited by such dramatic situations wherein someone is considering to directly and personally sacrifice fewer human beings for the higher good of the many lead to a cognitive conflict associated with the anterior cingulate cortex and can be overruled at least in some subjects and in some cases by cognitive control associated with the DLPFC (see Figure 1). The

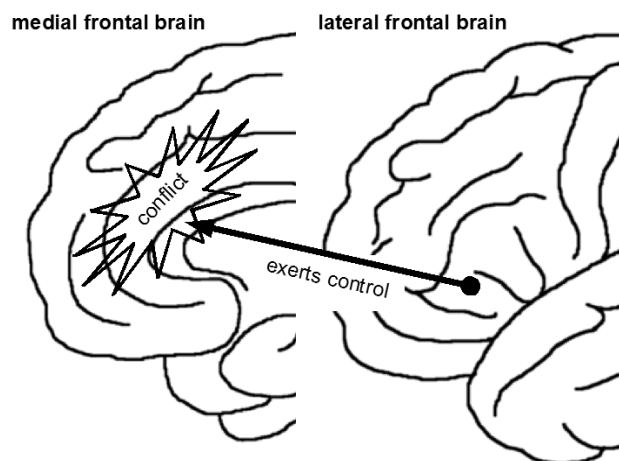


Figure 1. According to the Dual-Process Model, moral dilemmas cause an emotional conflict processed in the anterior cingulate cortex (left) that can, in some subjects, be overcome through cognitive control processes associated with the dorsolateral prefrontal cortex (DLPFC, right).

dual-process theory endorsed by Greene and colleagues thus describes moral judgment as the result of the competition between cognitive and emotional brain processes.

2.2. *The event-feature-emotion-complex model*

Proponents of another prominent theory of moral judgment have challenged this view (Moll and Oliveira-Souza 2007). Based on their own investigations of passive perception of morally salient pictures or short texts (Moll et al. 2002b, 2005b), simple moral judgments (Moll et al. 2001, 2002a), or charitable donation (Moll et al. 2006) within the fMRI scanner, Moll and colleagues endorse a network model of moral cognition, called the event-feature-emotion-complex (EFEC) that posits the integration of various brain mechanisms in moral perception, cognition, and action (Moll et al. 2005a).

In more detail, the prefrontal cortex is assumed to represent structured event knowledge, the temporal lobes social features, and limbic structures such as the amygdalae central motive states (for a selection, see Figure 2). Structured event knowledge consists in “context-dependent representations of events and event sequences” (Moll et al. 2005a:804), social features can either be perceptual (e.g. face, gaze, or body posture, posterior and superior part of the temporal lobe) or functional (e.g. functional features of social behaviours), and examples for central motive states are affiliative experience, hunger, or sexual arousal. Thus, whereas the idea behind Greene and colleagues’ model is that of conflict between emotion and cognition, Moll and colleagues’ theory is based on an integration of emotional, social, and cognitive (e.g. event knowledge) aspects.

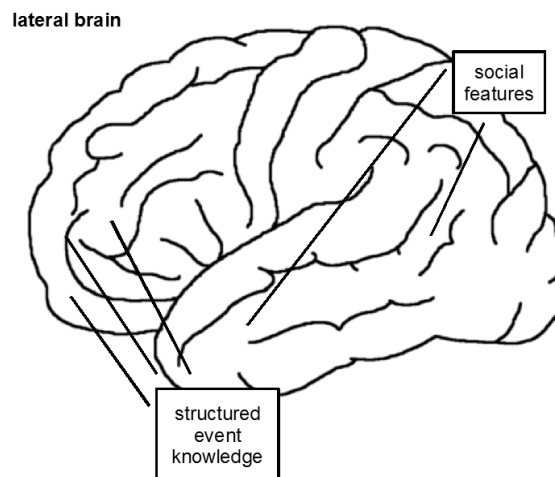


Figure 2. According to the event-feature-emotion-complex model, moral cognition integrates cognitive aspects represented in the prefrontal cortex, social features represented in the temporal lobes, and central motive states represented in limbic structures (not shown).

In summary, the EFEC framework attempts to unite a variety of findings from studies on moral judgment in particular and from research on brain processing in general. The complexity of the framework is a result of this all-encompassing approach.

2.3. The social intuitionist model

The third model originates in social-psychological and transcultural research that focused on the influence of disgust on moral judgment (Haidt, Koller, and Dias 1993, Haidt et al., 1997) and combines psychological, evolutionary and neuroscientific perspectives (Greene and Haidt 2002, Haidt 2007). Under the impression that disgust influences moral judgments (Wheatley and Haidt 2005, Schnall et al. 2008) and that people often cannot justify their moral decisions (Haidt 2001), Jonathan Haidt developed the Social Intuitionist Model (SIM) proposing the idea that intuition drives moral judgment. Moral intuition is considered as a kind of sudden and unconscious cognition incorporating moral emotions that directly cause moral judgment (Haidt 2001). Thus, neither moral intuition nor moral judgment is subject to cognitive control: When asked to judge a moral situation, people are deemed to decide intuitively without reasoning or deliberation. Hence, intuition automatically generates the judgment that is justified only afterward through post-hoc reasoning (Haidt and Kesebir 2010).

Furthermore, the model makes suggestions about the personal and social consequences of a given moral judgment by allowing for private reflection, social justification, and persuasion. This means that the moral subject herself as well as her social environment reasons upon the outcome of a moral judgment. People do reflect on their judgments after they are made and there is a tendency to arrange one's judgment with one's own actions and the expectations of other people. This is achieved by evaluation that can lead to a change in the judge's intuition or in her social environment. This altered intuition then proceeds to function in the automatic way described above (see Figure 3; Haidt 2001).

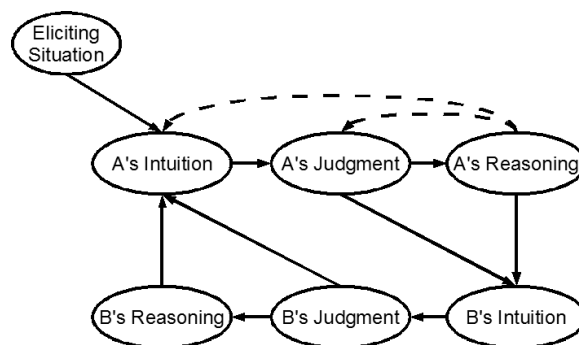


Figure 3. Although moral judgments are based on sudden and unconscious intuitions according to the social intuitionist model and only justified rationally after the decision is made, the model allows at least some reasoning, as well as the judgments and reactions of others to influence one's moral intuitions.

In summary, the SIM depicts moral judgment as the result of intuition, reasoning and social influences, with a primacy of intuition. This conception draws on a variety of findings from social psychology, e.g. studies on post-hoc reasoning (Nisbett and Wilson 1977), automaticity (Bargh and Chartrand 1999), and dual-process theory (Chaiken and Trope 1999). The theory is a reformulation of these findings aimed at describing and explaining moral judgments.

2.4. Discussion

Despite their differences, the three approaches share a common understanding of moral judgment: as framed and studied in the fashion of cognitive science. The approaches all provide models of the functioning of moral judgment in terms of distinctive network-modules with specific capacities that interact via excitation/inhibition or feedback/feed-forward mechanisms. This is an idiosyncratic way of describing and understanding moral judgment in terms of moral cognition (encompassing intuition and emotion). This understanding defines the common goal of the associated research, namely, to uncover the cognitive processes of moral judgment by identifying its relevant components and providing models of their interplay.

This modularized cognitive-processing view of morality shapes moral physiology in a certain way, particularly in combination with a historical debate in moral philosophy that is frequently referred to in the recent empirical literature (Greene 2007, Haidt 2001, Haidt and Kesebir 2010, Huebner, Dwyer, and Hauser 2009, Monin, Pizarro, and Beer 2007, Schnall et al. 2008), that is, the classical debate between David Hume and Immanuel Kant on the role of emotion and reasoning in moral judgment, but the topos of a competition between passion and reason in human actions is much older and can already be found in the Bible (Mt 26,40-41, Rom 7, 15). Experiments are nowadays carried out and/or interpreted in such a way as to confirm which has in another place been explicitly called the ‘Humean’, ‘Kantian’, or ‘Rawlsian’ model of moral judgment (Hauser 2006). Through laboratory manipulations of the emotional content of moral stimuli, researchers want to check whether moral judgment is subject to emotion and conclude accordingly that it is not entirely grounded in reason. In this endeavour we see a tendency for a view that might be called ‘moral essentialism’ – the quest to uncover what kind of cognitive processing human morality is ‘really’ grounded in.

Whether or not there is an ‘essence’ of moral judgment, we doubt that the logic of laboratory manipulations is sufficient to make this case. For example, when researchers can successfully demonstrate that variations of emotional content giving rise to dilemmatic conflict (Greene et al. 2001, 2004), disgust induced by hypnosis or a dirty environment (Wheatley and Haidt 2005, Schnall et al. 2008), differences in cognitive load or working memory capacity (Greene et al. 2008, Moore, Clark, and Kane 2008), or sleep deprivation (Killgore et al. 2007), to name just a few examples, all can influence moral judgment, this only shows that moral cognition – like probably every mental faculty – is amenable to a multitude of environmental and psychological influences. Likewise, developmentalists tried to

uncover developmental (Kohlberg and Puka 1994, Piaget 1932) or situationists' situational influences (Carpendale and Krebs, 1992) on moral cognition and behaviour. Yet, although they identified certain aspects that can influence human morality, this did not mean that moral judgment was 'essentially' a developmental or situational capacity.

Moral cognition and behaviour is not only subject to a multitude of physiological, psychological, and social aspects, but the processes suggested as central in the different models, such as cognitive control, dual processing, or post-hoc reasoning, have previously been related to other psychological functions (Botvinick et al. 2001, Chaiken and Trope 1999, Koechlin et al. 1999, Nisbett and Wilson 1977). Instead of reducing morality to only one kind of processing or to only one model, we defend a multidisciplinary and multimodal view on morality to which we return in section 4 after a theoretical analysis of the empirical work described so far.

3. Theoretical reflection

It goes without saying that understandings of research purporting to give answers to normative questions have to be based on a solid experimental and methodological basis. After all, if interpretations were so arbitrary or if data were so ambiguous that a multitude of different moral standpoints could be justified with them, there would be little justification for the defence of one particular stance in contrast to others. As expressed by Sperry above and as witnessed by thousands of studies since the officially proclaimed "Decade of the Brain" 1990–2000 (Bush 1990), the brain is believed to give many answers to psychological riddles. Greene and colleagues' endeavour to understand why a majority finds it acceptable in the dilemmas described above to throw the switch, taking into account the death of one, but unacceptable to push the stranger when the lives of the five workers are at stake, is a prime example (Greene et al. 2001).

Within the field of cognitive and behavioural neuroscience, neuroimaging has taken the lead (Friston 2009) and within neuroimaging, fMRI has rapidly become a dominant technique (Logothetis and Wandell 2004) with its many ways of investigating and visualizing brain function in such a way that it even has a (modestly) convincing effect on lay people (McCabe and Castel 2008, Weisberg et al. 2008). Although the number of yearly publications related to fMRI has exponentially grown since its inception in 1990 and already exceeds 2000 by far (Schleim 2011) and although ever more of its findings are covered in public media (Racine, Bar-Ilan, and Illes 2005), surprisingly little attention is paid to the brain's actual complexity and the method's limitations (Logothetis 2008, Racine, Bar-Ilan, and Illes 2006).

The intention of this section is to cover some of these aspects. This is particularly relevant to understanding the scope of the alleged normative implications of moral physiology as well as other kinds of practical applications of neuroimaging (Schleim and Rosier 2009) and thus also central to debates in neuroethics generally that are

based on the technological state of the art. We start out with the difference between the experimental and real life situations, what is frequently subsumed under the concepts of ‘external’ and ‘ecological validity’. After this discussion of validity, we turn to the operational definition of moral judgments and utilitarian decisions as employed in the influential studies of Greene and colleagues. We then continue with some basic aspects of brain neurobiology, anatomical localization, and inferring cognitive processes from brain activation. These issues might seem rather technical and remote from the questions of moral physiology and its normative implications, but they are actually central to the understanding of the experiments’ scope and the validation of the models described in section 2.

3.1. Experiment and real life

There are not only many *conceivable* ways to understand and investigate human morality, but also many different *actual* practices with which researchers approach the topic. We have briefly referred to the different kinds of moral dilemmas used by Greene and colleagues (2001, 2004) which are related to scholarly debates on moral issues but in most cases also very abstract. For example, in one case the subjects are asked to imagine that their family unknowingly camped on the sacred ground of a tribe, thereby desecrating it, and the only way to prevent the upset clan people from killing the whole family is to sacrifice the life of one of their children with their own hands (Greene et al. 2001, 2004). One of the authors of this paper has used short stories adapted from actual moral and political cases and asked his experimental subjects to judge whether a certain action is right from a moral point of view (Schleim et al. 2011).

Hauke Heekeren and colleagues let their subjects judge simple sentences such as “A uses public transport without paying” (Heekeren et al. 2003, 2005, Prehn et al. 2008), allowing them a stricter experimental control than Greene and colleagues or Schleim and colleagues but at loss of complexity of the moral issues. Moll and colleagues used different classes of pictures based on subjects’ evaluation of their moral content on a scale from one to ten. ‘Moral content’ was explained as including “actions which you consider to be commendable or regrettable, fair or unfair, right or wrong, good or evil, or situations that evoke a sense of friendship, betrayal, pity or care for others, humiliation, gratitude, or indignation” (Moll et al. 2002b:2731). As a consequence, their moral category had an average rating of 7.13, but ‘neutral’ and ‘pleasant’ pictures were not completely devoid of moral content with mean ratings of 3.73 and 4.5 respectively.

It is not our intention to prescribe *the* correct way to investigate morality. Yet we want to briefly discuss how far these already different and varying complex ways of investigating moral cognition and behaviour reflect actual moral situations in real life, such as when a couple is pondering to abort the life of an unwanted or unhealthy foetus, or a political committee has to decide to spend limited resources on the protection of people in one place at the expense of the safety of others (Turiel 2010). Both of these examples, we can easily imagine, are not only subject to societal and technological contexts, but can also involve interactions with and

decisions of people in different personal and institutional roles. No experimental setting within moral physiology so far has been able to include these wider contexts and it is unclear whether any future experiment will do so. Simply letting two or more people all lying in brain scanners interact with each other, a technology sometimes referred to as ‘hyperscanning’ (Casebeer 2003), will not solve this issue.

These considerations limit the external and ecological validity of the respective experiments meaning that they constrain generalizations from laboratory to real life moral judgments. This is a common issue in all experimental sciences and is relevant to their practical applications. In moral physiology, applications consist in drawing conclusions on the normative level. The laboratory settings employed so far rather resemble situations in which people are pondering moral issues from a ‘what if’ perspective but where they are not necessarily personally involved, because their decisions have no or little consequences for themselves, but even then the experiments are devoid of social feedback, that is, those loops which are so central to Haidt’s SIM (see section 2.3).

3.2. *Of ‘utilitarian’ and ‘moral personal’ dilemmas*

The most important finding of Greene and colleagues and a central support for their dual-processing model of moral judgment was an increased activation of the DLPFC (though see section 3.3) associated with cognitive control when subjects were making ‘utilitarian’ choices, in their words, “judgments that maximize aggregate welfare (e.g. by sacrificing one life in order to save five others)” (Greene et al. 2004:390). In combination with the previous findings of the ‘emotional interference effect’ that subjects choosing the ‘utilitarian’ option took longer and that ‘moral personal’ dilemmas engage emotion (Greene et al. 2001), they and Peter Singer (2005) undermined counter-utilitarian intuitions as irrational. In Greene’s own words, it is “the secret joke of Kant’s soul” (2007:35) that the allegedly rational Kantian moral philosophy and particularly its absolute prohibitions are rather based on emotion than reason.

These far-reaching conclusions call for some reflection on how Greene and colleagues relate their neuroscientific findings to moral-philosophical categories such as utilitarianism and (Kantian) deontology. Their final fMRI contrast yielding the DLPFC activation is based on a comparison between those difficult (i.e. cases in which subjects need more time to answer) ‘moral personal’ cases in which subjects choose ‘yes, appropriate’ in response to the sacrificing option instead of ‘no, inappropriate’. However, even if the average utilitarian indeed chose these sacrificing options, something only assumed but never demonstrated by Greene and colleagues, this does not make these options necessarily utilitarian (Schleim 2008). One can illustrate this with reference to another ‘moral personal’ dilemma, where a bleeding hiker lies next to a road and the sacrifice to save his life consists in ruining the expensive leather upholstery of one’s car. Hardly a moral theory deserving this name would contradict the sacrificing option. That is, proponents of

different views, be they utilitarians, deontologists, virtue theorists, or others, would equally endorse this option.

A more systematic analysis based on responses of moral philosophers at the Oxford University has yielded the result that only 45% of the ‘moral impersonal’ and 48% of the ‘moral personal’ cases actually allowed a choice between utilitarian and non-utilitarian options (Kahane and Shackel 2008). A later study based on the original criteria of Greene and colleagues controlled for additional features within the stimulus material, such as whether one’s own life will be affected if someone does refrain from choosing the sacrifice (Moore, Clark, and Kane 2008). Remember that the dilemma due to the desecration of the tribe’s sacred ground described above comprised the death of the whole family, including the parent (i.e. the experimental subject) facing the alternative to kill one’s own child. Yet, the maximization of the common good favoured by utilitarians usually does not place one’s own life above that of others but was historically directed against moral egoism (Sidgwick 1907). The study of Moore and colleagues has shown that these non-utilitarian aspects present but not controlled in Greene and colleagues’ stimulus material does matter behaviorally.

Furthermore, with their improved stimulus material, Moore and colleagues could not replicate the “emotional interference effect” (Moore, Clark, and Kane 2008). Moreover, the group of McGuire and colleagues who re-analyzed the reaction-time data of Greene and colleagues has shown that the ‘moral impersonal’ and ‘personal’ categories were not homogeneous and that the ‘emotional interference effect’ was rather driven by a quick refusal of sacrificing options than their slower endorsement as hypothesized by the dual-process model (McGuire et al. 2009). When they checked for these aspects, the ‘emotional interference effect’ disappeared.

More conceptual and methodological caveats have been put forward against the studies by Greene and colleagues and their normative conclusions (Berker 2009, Kahane and Shackel 2010, Kamm 2009, Schleim 2008, 2011) and Greene himself has given up his original categorization (Greene 2007). Yet, the distinction between ‘moral impersonal’ and ‘personal’ dilemmas, as well as the interpretation of ‘utilitarian’ choices has been very influential as witnessed by more than 600 citations of the two studies and an endorsement of the experimental design by many other groups in their own research until today (e.g. Ciaramelli et al. 2007, Glenn, Raine, and Schug 2009, Koenigs et al. 2007, Crockett et al. 2010). After discussing the issues of validity and operational definitions in moral physiology, we turn now to broader aspects related to fMRI as a method to uncover the neurobiological underpinnings of psychological processes.

3.3. Blood flow and brain activation

Lay people as well as scholars from other disciplines might take it for granted that modern methods of neuroscience directly investigate brain activation. Even if this were the case, there are many different electrochemical, spatial, and temporal properties of brain processes and each method is only related to a subset of these. The huge success of fMRI is based on a good compromise between spatial and

temporal accuracy and the general tolerability of even high magnetic fields comprising the methodological basis. Yet, every single data point collected by an fMRI device, called a ‘voxel’, with standard parameters still contains a whole cosmos of brain processes in itself, containing approximately 27 mm^3 of tissue including 540,000 to 2.7 million neurons with 11 to 27 billion synapses, more than 10 km of dendrites and 100 km of axons (Logothetis 2008) besides other kinds of cells. *One* value representing an aggregate of the whole voxel is usually recorded every two seconds and the whole brain represented by several tens of thousands of these units. The areas of activation reported in fMRI studies usually comprise ten to several hundreds of voxels. Technological progress will improve some of these aspects, but ultimately safety considerations and biological tolerability will pose a limit.

More relevant than the technological limitation is the actual neurobiology behind fMRI. The method is based on the fact that different properties of blood oxygenation primarily based on changes of blood flow have different magnetic properties (Heeger and Ress 2002, Logothetis and Wandell 2004, Raichle and Mintun 2006). The idea that blood flow represents brain activation has a long history in neuroscience (Mosso 1881) and recent research has indeed found correlations between the fMRI signal and neural processing in animals as measured with electrophysiological instruments (Logothetis et al. 2001). Yet, blood flow and brain activation are not identical and many experiments dissociated them many experiments dissociated them (Schummers, Yu, and Sur 2008, Sirotin and Das 2009) and the strength of the association differs between different brain areas (Ekstrom 2010). Notwithstanding, fMRI results are frequently described as ‘neural activation’, ‘neural processes’, ‘neural correlates’, and so on (Schleim and Rosier 2009), which they are not necessarily. The interpretation of the signal and thus the understanding of what fMRI data actually show us is still a question of ongoing basic research. These considerations do not suggest that the many published studies are uninteresting but rather that they are to be interpreted with caution.

3.4. Anatomical localization

Unlike other methods taking account of network properties of the brain, fMRI essentially is a localizational technique, that is, a means to pinpoint signal differences in a three-dimensional space. In order to ultimately interpret these differences in psychological terms, two kinds of procedures are necessary (see also section 3.5): First, generalizations to whole populations require the measurement of several persons and the transformation of their individual brains into a standard space, second, the localized places must be assigned anatomical labels.

The upshot of the first procedure is that brain localization in groups is essentially probabilistic (Zilles and Amunts 2010), since every brain is structurally different. There are some areas with higher and some with lower between-subject variability, and some amount of variability also comes with normal ageing. Even in a highly homogeneous group of 2500 young and healthy men applying for military air service, Frank Weber and Heinz Knopf (2006) reported norm-deviations and abnormalities in about 25% with the naked eye based on structural

MRI investigations. This does not make brain localization impossible, but it emphasizes that each individual brain differs from the standard space into which its signals are finally transformed.

But also in a standard space the localizations are not yet meaningful. To assign anatomical labels to the places in three-dimensional space, different kinds of templates are used, for example the already more than 100-year-old map of Korbinian Brodmann based on microscopic investigations of brain tissue, the map of a dissected brain of a 60-year-old French woman in a so-called Talairach space, or the MNI-atlas based on 305 anatomical MRI images of young, right-handed, North-American and mostly male healthy volunteers. The existence of standard spaces makes localized brain signals comparable, yet some amount of ambiguity remains.

Combining different kind of standards or levels of individuation such as describing brain areas on the very coarse-grained levels of whole lobes or spatially oriented subsections thereof can suggest different interpretations. For example, the primary activation associated with ‘utilitarian’ judgment reported as DLPFC by Greene and colleagues (2004), that is, the place very much in front (‘prefrontal’), towards the top (‘dorsal’ as opposed to ‘ventral’) and rather on the side (‘lateral’ as opposed to ‘medial’), was localized in the medial frontopolar cortex by Moll and de Oliveira-Souza (2007) instead, although ‘lateral’ and ‘medial’ are mutually exclusive labels. On a broader scale, Tonio Ball and colleagues used a recently developed microscopic probabilistic atlas of some limbic structures to re-analyze 335 localizations reported as amygdala activations in the period from 2000 till 2008 (Ball et al. 2009). They found that only 49% of them belonged to this area with a high probability (>80%), in 15% of the cases the probability was 0%. Respective maps combining macroscopic and microscopic features for the whole brain in different populations, taking into account variability due to gender, ethnicity, age and further features, are still a matter of basic research. The following section will explain further why this is important.

3.5. Inferring cognitive processes

If we follow the stepwise logic of this section, we are now finally at the stage of assigning a psychological function to a spatially localized difference in a blood-flow measurement associated with a certain experimental condition. As has been outlined by Russell Poldrack, this inference is usually carried out in three steps. The first step is the said result of the localization procedure associated with a particular task. In the second step, the identified brain area is compared to the body of known literature, especially other studies which found the same area when a certain cognitive process was (putatively) present. In the third and final step, the former two are combined to conclude that the activity of the brain area in the present study shows engagement of that cognitive process (Poldrack 2006).

It is immediately apparent why this inference is, logically speaking, valid only under a special condition, namely that of a 1:1- or n:1-mapping between brain areas and cognitive functions. As soon as the respective brain area has been

associated with more than one function (i.e. 1:n or n:m), it cannot (at least not logically) be inferred that activation of that brain area shows engagement of one particular cognitive process in contrast to others. Although most brain areas are associated with a multitude of functions, even such paradigmatic cases of specialization as the language-related Broca's area (Anderson 2010), Poldrack (2006) himself proposed a statistical method based on Bayes' theorem to take account of the varying amount of functional specialization of brain areas and consequently to assign a level of probability or certainty to the inference, though it is hardly used in practice.

It has been noted that some researchers instead tend to interpret their findings in terms of their own domains, that is, differences in activation "were usually attributed to episodic memory processes in episodic memory studies, visuo-spatial processes in visuo-spatial studies and so on" (Cavanna and Trimble 2006:579) which implies some amount of circularity since that presumes already what the studies are designed to find out. Another strategy consists in including other measures common in psychology, such as behavioural pilot studies, reaction times, peripheral physiology, questionnaires, interviews, and so on. Although this undermines the original contribution of neuroimaging to understanding the human mind, it also takes account of the method's limitations and the natural structure and functioning of the human brain. In summary, this subsection in combination with the previous one demonstrates that independent of the primary statistical method used to test for significant differences in the fMRI signal, additional dimensions of probability are added at the stages of anatomical localization and inferring cognitive processes.

4. Implications and challenges

We started out with many instances of alleged philosophical implications of moral physiology in the introduction and confronted these claims with sobering methodological and theoretical reflections in the previous section. The most far-reaching consequence would certainly be to 'read' the right moral answers from people's brains. For the sake of the argument, we now assume that there was indeed a sufficiently replicated body of moral-physiological research uncovering the cognitive processes of moral judgment. Let us assume, first, that some particular kind of moral judgment J characteristic of a certain moral theory T was indeed based on cognitive processing P. While this finding would be empirically interesting, we do not see how it could *taken by itself* have any normative force. Similar to Moore's open question argument (Moore 1903), one might ask of this finding: "Is it morally right that J is based on P?" We think that some additional argument would be necessary to answer this question, an argument that could not ultimately be answered by another brain scan, of whose result the same normative question could be asked. Yet, the finding that J is based on P might be normatively relevant, for example, if T contained the view that J should not be based on P but

on Q, or if a meta-ethical argument implied that proper moral judgments must not be based on Ps but on Qs. In both cases, however, the normative force of the finding that J is based on P essentially requires the T-component or meta-ethical argument that this is not morally right. As a response, one might give up the T-component, T completely or provide a sound counterargument.

Let us briefly turn to the idea proposed by Casebeer and Churchland that moral physiology “may allow us to eliminate certain moral theories as being psychologically and neurobiologically unrealistic” (Casebeer and Churchland 2003:171). According to the principle that *ought implies can* one might feel inclined to conclude that a moral theory T were morally implausible if it systematically required judgments of kind J that were impossible to process by the average person due to some psychological or neurobiological constraint. Leaving aside that this inclination presumed the additional principle, one might ask further what such a finding could mean. Apparently at least the person who developed T or now tries to test it was able to process that J, according to T, is right. So one might argue that even if the average person is unable to process this, T might still function as a general guideline.

But perhaps being psychologically and neurobiologically unrealistic means that J is not impossible to process as a judgment, but impossible to carry out. Obviously cases happen in which people perform non-J, although they believe that to the contrary J is right, something one might consider a pity dilemma of the human condition already described in many examples of philosophy, history, and the literature. However, we think that the claim that J is, on psychological and neurobiological grounds, impossible to carry out, seems to be a very strong one. Particularly given our knowledge on brain development, the effects of brain training, enhancement, and neuroplasticity, it appears to be very difficult to claim that in principle, by psychological or neurobiological necessity it is impossible to carry out J, so difficult that we think that the onus of proof is upon those who are seeking this kind of ‘elimination’.

The upshot of this section then is that moral physiology might have normative implications but that it is implausible to believe that moral answers might be ‘read’ directly from human brains without the necessity of additional normative premises or the possibility of counterarguments (Schleim 2011). Section 3 also suggests that empirical issues are none the less controversially debated than normative issues and theories. The idea that moral philosophers might become superfluous due to progress in neuroscience seems unwarranted and is indeed contradicted by the increasing number of philosophical publications reflecting on moral physiology. The multi-disciplinary challenge, on the contrary, might consist in remembering that not all questions can be answered by brain research alone, that particularly the dimension of human morality extends far beyond the brain and encompasses a multitude of cultural, environmental, institutional, social, and technological contexts, that instead of searching for a moral ‘essence’ and reducing morality to a single model, emotion, reason, cognitive control, and intuition all might play different roles in different kinds of situations and subjects, that given

the amount of open conceptual, empirical, methodological, and theoretical questions it might be wise to avoid ‘brain overclaim’, to use the notion coined by Stephen Morse in contexts of penal law (Morse 2006), that is, a tendency to claim that neuroscientific research has normative implications it apparently does not have, or, in different terms, to avoid ‘neuro-realism’, ‘neuro-essentialism’, and ‘neuro-policy’, as coined by Eric Racine and colleagues (Racine, Bar-Ilan, and Illes 2005).

5. Conclusion

The discussion of moral physiology with its empirical, cultural, and theoretical aspects, demonstrates that there remains a variety of challenges to investigate human morality from the perspective of many disciplines. While neurosciences have contributed an unprecedented amount of new knowledge to the functioning of the human brain, it is not evident how a neuroscientific result could exert a normative force just by itself. By contrast, such results are also in need of interpretation and the conclusions drawn from them also require justification, including normative considerations where normative questions are involved.

We summarized influential models of contemporary moral physiology, addressed some of the central experiments and their interpretations, and discussed empirical, methodological, and theoretical aspects of one of the leading techniques within the prospering field of behavioural and cognitive neuroscience, namely, fMRI. Particularly this latter discussion is not only relevant to moral physiology, but the broader context of neuroethics as well – insofar as the ethical implications of progress in and applications of neurotechnology, particularly inasmuch as they are based on fMRI, are involved. The widely acknowledged complexity of the human brain indeed fosters the success of neuroimaging and promises to yield even more new knowledge in the future; yet, it is just the limiting aspect of complexity reminding us that every measurement provides only a partial perspective on a particularly selected aspect of reality. Finally, the complexity of human beings and their societies encompasses but ultimately surpasses that of individual brains.

Address:

Stephan Schleim and Felix Schirmann
Theory and History of Psychology
Faculty of Behavioral and Social Sciences
University of Groningen
Grote Kruisstraat 2/1
9712 TS Groningen
The Netherlands

E-mail: s.schleim@rug.nl

Tel.: +31 (0)50 363 6244

Acknowledgments

We would like to thank Professors Birnbacher, Gethmann, Kahane, Kleingeld, Metzinger, Nobuhara, Schöne-Seifert, Stephan, and Walter for the possibility to present earlier drafts of this work at their conferences or colloquia. We would also like to acknowledge the many important comments we received on these occasions by the respective audiences. This paper was supported by the grant “Intuition and Emotion in Moral Decision-Making: Empirical Research and Normative Implications” by the Volkswagen Foundation, Az. II/85 063.

References

- Anderson, M. L. (2010) “Neural re-use as a fundamental organizational principle of the brain”. *Behavioral and Brain Sciences* 33, 245–313.
- Ball, T., J. Derix, J. Wentlandt, B. Wieckhorst, O. Speck, A. Schulze-Bonhage, et al. (2009) “Anatomical specificity of functional amygdala imaging of responses to stimuli with positive and negative emotional valence”. *Journal of Neuroscientific Methods* 180, 1, 57–70.
- Bargh, J. A. and T. L. Chartrand (1999) “The unbearable automaticity of being”. *American Psychologist* 54, 7, 462–479.
- Batson, C. D., B. D. Duncan, P. Ackerman, T. Buckley, and K. Birch (1981) “Is empathic emotion a source of altruistic motivation”. *Journal of Personality and Social Psychology* 40, 2, 290–302.
- Berker, S. (2009) “The normative insignificance of neuroscience”. *Philosophy and Public Affairs* 37, 4, 293–329.
- Botvinick, M. M., T. S. Braver, D. M. Barch, C. S. Carter, and J. D. Cohen (2001) “Conflict monitoring and cognitive control”. *Psychological Review* 108, 3, 624–652.
- Buckholz, J. W., C. L. Asplund, P. E. Dux, D. H. Zald, J. C. Gore, O. D. Jones, et al. (2008) “The neural correlates of third-party punishment”. *Neuron* 60, 5, 930–940.
- Bush, G. (1990) “Presidential Proclamation 6158”. Retrieved January 3, 2011, from <http://www.loc.gov/loc/brain/proclaim.html>
- Carpendale, J. I. M. and D. L. Krebs, (1992) “Situational variation in moral judgment – in a stage or on a stage”. *Journal of Youth and Adolescence* 21, 2, 203–224.
- Casebeer, W. D. (2003) “Moral cognition and its neural constituents”. *Nature Reviews Neuroscience* 4, 10, 840–846.
- Casebeer, W. D. and P. S. Churchland (2003) “The neural mechanisms of moral cognition: a multiple-aspect approach to moral judgment and decision-making”. *Biology and Philosophy* 18, 1, 169–194.
- Cavanna, A. E. and M. R. Trimble (2006) “The precuneus: a review of its functional anatomy and behavioural correlates”. *Brain* 129, 3, 564–583.
- Chaiken, S. and Y. Trope (1999) *Dual-process theories in social psychology*. New York: Guilford Press.
- Ciaramelli, E., M. Muccioli, E. Làdavas, and G. di Pellegrino (2007) “Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex”. *Social Cognitive and Affective Neuroscience* 2, 84–92.
- Crockett, M. J., L. Clark, M. D. Hauser, and T. W. Robbins (2010) “Serotonin selectively influences moral judgment and behavior through effects on harm aversion”. *Proceedings of the National Academy of Sciences of the United States of America* 107, 40, 17433–17438.
- Ekstrom, A. (2010) “How and when the fMRI BOLD signal relates to underlying neural activity: the danger in dissociation”. *Brain Research Review* 62, 2, 233–244.
- Farah, M. J. (2010) *Neuroethics: an introduction with readings*. Cambridge, Mass.: MIT Press.

- Freud, S. (1930) *Das Unbehagen in der Kultur*. Wien: Internationaler Psychoanalytischer Verlag.
- Friston, K. J. (2009) "Modalities, modes, and models in functional neuroimaging". *Science* 326, 5951, 399–403.
- Gazzaniga, M. S. (2005) *The ethical brain*. New York and Washington, D.C.: DANA Press.
- Giordano, J. J. and B. Gordijn (2010) *Scientific and philosophical perspectives in neuroethics*. Cambridge and New York: Cambridge University Press.
- Glenn, A. L., A. Raine, and R. A. Schug (2009) "The neural correlates of moral decision-making in psychopathy". *Molecular Psychiatry* 14, 1, 5–6.
- Greene, J. and J. Haidt (2002) "How (and where) does moral judgment work?". *Trends in Cognitive Sciences* 6, 12, 517–523.
- Greene, J. D. (2007) "The secret joke of Kant's soul". In *Moral psychology. The neuroscience of morality: emotion, brain disorders, and development*. Vol. 3, 35–79. W. Sinnott-Armstrong, ed. Cambridge, MA: MIT Press.
- Greene, J. D., S. A. Morelli, K. Lowenberg, L. E. Nystrom, and J. D. Cohen (2008) "Cognitive load selectively interferes with utilitarian moral judgment". *Cognition* 107, 3, 1144–1154.
- Greene, J. D., L. E. Nystrom, A. D. Engell, J. M. Darley, and J. D. Cohen (2004) "The neural bases of cognitive conflict and control in moral judgment". *Neuron* 44, 2, 389–400.
- Greene, J. D., R. B. Sommerville, L. E. Nystrom, J. M. Darley, and J. D. Cohen (2001) "An fMRI investigation of emotional engagement in moral judgment". *Science* 293, 5537, 2105–2108.
- Haidt, J. (2001) "The emotional dog and its rational tail: a social intuitionist approach to moral judgment". *Psychological Review* 108, 4, 814–834.
- Haidt, J. (2007) "The new synthesis in moral psychology". *Science* 316, 5827, 998–1002.
- Haidt, J. and S. Kesebir (2010) "Morality". In *Handbook of Social Psychology*. 5th ed., 797–832. S. Fiske, D. Gilbert, and G. Lindzey, eds. Hoboken, NJ: Wiley.
- Haidt, J., S. H. Koller, and M. G. Dias (1993) "Affect, culture, and morality, or is it wrong to eat your dog". *Journal of Personality and Social Psychology* 65, 4, 613–628.
- Haidt, J., P. Rozin, C. R. McCauley, and S. Imada (1997) "Body, psyche, and culture: the relationship between disgust and morality". *Psychology and Developing Societies* 9, 107–131.
- Hauser, M. D. (2006) "The liver and the moral organ". *Social Cognitive and Affective Neurosciences* 1, 3, 214–220.
- Heeger, D. J. and D. Ress (2002) "What does fMRI tell us about neuronal activity?". *Nature Reviews Neuroscience* 3, 2, 142–151.
- Heekeren, H. R., I. Wartenburger, H. Schmidt, K. Prehn, H. P. Schwintowski, and A. Villringer (2005) "Influence of bodily harm on neural correlates of semantic and moral decision-making". *Neuroimage* 24, 3, 887–897.
- Heekeren, H. R., I. Wartenburger, H. Schmidt, H. P. Schwintowski, and A. Villringer (2003) "An fMRI study of simple ethical decision-making". *Neuroreport* 14, 9, 1215–1219.
- Helmuth, L. (2001) "Cognitive neuroscience – moral reasoning relies on emotion". *Science* 293, 5537, 1971–1972.
- Huebner, B., S. Dwyer, and M. Hauser (2009) "The role of emotion in moral psychology". *Trends in Cognitive Sciences* 13, 1, 1–6.
- Illes, J. (2006) *Neuroethics: defining the issues in theory, practice, and policy*. Oxford and New York: Oxford University Press.
- Kahane, G. and N. Shackel (2008) "Do abnormal responses show utilitarian bias?". *Nature* 452, E5.
- Kahane, G. and N. Shackel (2010) "Methodological issues in the neuroscience of moral judgement". *Mind and Language* 25, 5, 561–582.
- Kamm, F. M. (2009) "Neuroscience and moral reasoning: a note on recent research". *Philosophy and Public Affairs* 37, 4, 330–345.
- Killgore, W. D. S., D. B. Killgore, L. M. Day, C. Li, G. H. Kamimori, and T. J. Balkin (2007) "The effects of 53 hours of sleep deprivation on moral judgment". *Sleep* 30, 3, 345–352.
- Koechlin, E., G. Basso, P. Pietrini, S. Panzer, and J. Grafman (1999) "The role of the anterior prefrontal cortex in human cognition". *Nature* 399, 6732, 148–151.
- Koenigs, M., L. Young, R. Adolphs, D. Tranel, F. Cushman, M. Hauser, et al. (2007) "Damage to the prefrontal cortex increases utilitarian moral judgements". *Nature* 446, 7138, 908–911.

- Kohlberg, L. and B. Puka (1994) *Kohlberg's original study of moral development*. New York: Garland.
- Levy, N. (2007) *Neuroethics*. Cambridge and New York: Cambridge University Press.
- Logothetis, N. K. (2008) "What we can do and what we cannot do with fMRI". *Nature* 453, 7197, 869–878.
- Logothetis, N. K., J. Pauls, M. Augath, T. Trinath, and A. Oeltermann (2001) "Neurophysiological investigation of the basis of the fMRI signal". *Nature* 412, 6843, 150–157.
- Logothetis, N. K. and B. A. Wandell (2004) "Interpreting the BOLD signal". *Annual Review of Physiology* 66, 735–769.
- McCabe, D. P. and A. D. Castel (2008) "Seeing is believing: the effect of brain images on judgments of scientific reasoning". *Cognition* 107, 1, 343–352.
- McGuire, J., R. Langdon, M. Coltheart, and C. Mackenzie (2009) "A reanalysis of the personal/impersonal distinction in moral psychology research". *Journal of Experimental Social Psychology* 45, 3, 577–580.
- Moll, J. and R. de Oliveira-Souza (2007) "Moral judgments, emotions and the utilitarian brain". *Trends in Cognitive Sciences* 11, 8, 319–321.
- Moll, J., R. de Oliveira-Souza, I. E. Bramati, and J. Grafman (2002a) "Functional networks in emotional moral and nonmoral social judgments". *Neuroimage* 16, 3, 696–703.
- Moll, J., R. de Oliveira-Souza, P. J. Eslinger, I. E. Bramati, J. Mourao-Miranda, P. A. Andreiuolo, et al. (2002b) "The neural correlates of moral sensitivity: a functional magnetic resonance imaging investigation of basic and moral emotions". *Journal of Neuroscience* 22, 7, 2730–2736.
- Moll, J., R. de Oliveira-Souza, F. T. Moll, F. A. Ignacio, I. E. Bramati, E. M. Caparelli-Daquer, et al. (2005b) "The moral affiliations of disgust – a functional MRI study". *Cognitive and Behavioral Neurology* 18, 1, 68–78.
- Moll, J., P. J. Eslinger, and R. de Oliveira-Souza (2001) "Frontopolar and anterior temporal cortex activation in a moral judgment task – preliminary functional MRI results in normal subjects". *Arquivos de Neuro-Psiquiatria* 59, 3B, 657–664.
- Moll, J., F. Krueger, R. Zahn, M. Pardini, R. de Oliveira-Souza, and J. Grafman (2006) "Human fronto-mesolimbic networks guide decisions about charitable donation". *Proceedings of the National Academy of Sciences of the United States of America* 103, 42, 15623–15628.
- Moll, J., R. Zahn, R. de Oliveira-Souza, F. Krueger, and J. Grafman (2005a) "The neural basis of human moral cognition". *Nature Reviews Neuroscience* 6, 10, 799–809.
- Monin, B., D. A. Pizarro, and J. S. Beer (2007) "Reason and emotion in moral judgment: different prototypes lead to different theories". In *Do emotions help or hurt decision making? A hedgefoxian perspective*, 219–244. K. D. Vohs, R. F. Baumeister, and G. Loewenstein, eds. New York: Russell Sage Foundation.
- Moore, A. B., B. A. Clark, and M. J. Kane (2008) "Who shall not kill? Individual differences in working memory capacity, executive control, and moral judgment". *Psychological Science* 19, 6, 549–557.
- Moore, G. E. (1903) *Principia ethica*. Cambridge: Cambridge University Press.
- Morse, S. J. (2006) "Brain overclaim syndrome and criminal responsibility: a diagnostic note". *Ohio State Journal of Criminal Law* 3, 397–412.
- Mosso, A. (1881) *Ueber den Kreislauf des Blutes im Menschlichen Gehirn. Untersuchungen*. Leipzig.
- Nadelhoffer, T., E. A. Nahmias, and S. Nichols (2010) *Moral psychology: historical and contemporary readings*. Malden, MA: Wiley-Blackwell.
- Nagel, S. K. (2010) *Ethics and the neurosciences: ethical and social consequences of neuroscientific progress*. Paderborn: Mentis.
- Nisbett, R. E. and T. D. Wilson (1977) "Telling more than we can know – verbal reports on mental processes". *Psychological Review* 84, 3 231–259.
- Piaget, J. (1932) *Le jugement moral chez l'enfant*. Paris: Librairie Fèlix Alcan.
- Poldrack, R. A. (2006) "Can cognitive processes be inferred from neuroimaging data?". *Trends in Cognitive Sciences* 10, 2, 59–63.

- Prehn, K., I. Wartenburger, K. Meriau, C. Scheibe, O. R. Goodenough, A. Villringer, et al. (2008) "Individual differences in moral judgment competence influence neural correlates of socio-normative judgments". *Social Cognitive and Affective Neuroscience* 3, 1, 33–46.
- Racine, E. (2010) *Pragmatic neuroethics: improving treatment and understanding of the mind-brain*. Cambridge, Mass.: MIT Press.
- Racine, E., O. Bar-Ilan, and J. Illes (2005) "fMRI in the public eye". *Nature Reviews Neuroscience* 6, 2, 159–164.
- Racine, E., O. Bar-Ilan, and J. Illes (2006) "Brain imaging – a decade of coverage in the print media". *Science Communication* 28, 1, 122–143.
- Raichle, M. E. and M. A. Mintun (2006) "Brain work and brain imaging". *Annual Review of Neuroscience* 29, 449–476.
- Roskies, A. (2002) "Neuroethics for the new millennium". *Neuron* 35, 1, 21–23.
- Schleim, S. (2008b) "Moral physiology, its limitations and philosophical implications". *Jahrbuch für Wissenschaft und Ethik* 13, 51–80.
- Schleim, S. (2011) *Die Neurogesellschaft: Wie die Hirnforschung Recht und Moral herausfordert*. Hannover: Heise Verlag.
- Schleim, S. and J. P. Roiser (2009) "fMRI in translation: the challenges facing real-world applications". *Frontiers in Human Neuroscience* 3, 63, 1–7.
- Schleim, S., T. M. Spranger, S. Erk, and H. Walter (2011) "From moral to legal judgment: the influence of normative context in lawyers and other academics". *Social Cognitive and Affective Neuroscience* 6, 48–57.
- Schnall, S., J. Haidt, G. L. Clore, and A. H. Jordan (2008) "Disgust as embodied moral judgment". *Personality and Social Psychology Bulletin* 34, 8, 1096–1109.
- Schummers, J., H. B. Yu, and M. Sur (2008) "Tuned responses of astrocytes and their influence on hemodynamic signals in the visual cortex". *Science* 320, 5883, 1638–1643.
- Sidgwick, H. (1907) *The method of ethics*. 7th ed. London: Macmillan and Co.
- Singer, P. (2005) "Ethics and intuitions". *Journal of Ethics* 9, 331–352.
- Sirotin, Y. B. and A. Das (2009) "Anticipatory haemodynamic signals in sensory cortex not predicted by local neuronal activity". *Nature* 457, 7228, 475–476.
- Sperry, R. W. (1981a) "Nobel lecture: some effects of disconnecting the cerebral hemispheres". Retrieved January 3 2011, from http://nobelprize.org/nobel_prizes/medicine/laureates/1981/sperry-lecture.html
- Sperry, R. W. (1981b) "Changing priorities". *Annual Review of Neuroscience* 4, 1–15.
- Thomson, J. J. (1985) "The Trolley problem". *The Yale Law Journal* 94, 6, 1395–1415.
- Turiel, E. (2010) "Snap judgment? Not so fast: thought, reasoning, and choice as psychological realities". *Human Development* 53, 3, 105–109.
- Unger, P. (1996) *Living high and letting die: our illusion of innocence*. New York: Oxford University Press.
- Weber, F. and H. Knopf (2006) "Incidental findings in magnetic resonance imaging of the brains of healthy young men". *Journal of Neurological Sciences* 240, 1–2, 81–84.
- Weisberg, D. S., F. C. Keil, J. Goodstein, E. Rawson, and J. R. Gray (2008) "The seductive allure of neuroscience explanations". *Journal of Cognitive Neurosciences* 20, 3, 470–477.
- Wheatley, T. and J. Haidt (2005) "Hypnotic disgust makes moral judgments more severe". *Psychological Science* 16, 10, 780–784.
- Young, L., F. Cushman, M. Hauser, and R. Saxe (2007) "The neural basis of the interaction between theory of mind and moral judgment". *Proceedings of the National Academy of Sciences of the United States of America* 104, 20, 8235–8240.
- Zilles, K. and K. Amunts (2010) "Centenary of Brodmann's map – conception and fate". *Nature Reviews Neuroscience* 11, 2, 139–145.