MORAL RATIONALISM UNDER EMPIRICAL ASSESSMENT

By
Marko Jurjako

Submitted to
Central European University
Department of Philosophy

In partial fulfilment of the requirements for the degree of Masters of Arts in Philosophy

Supervisor: Professor Christophe Heintz

Budapest, Hungary
2010

ABSTRACT

In this thesis I explore the impact and arguments that were based on recent discoveries in empirical moral psychology on the explicit and implicit ideas of philosophical moral rationalism. I follow the philosophical literature in differentiating conceptual, psychological and justificatory moral rationalism and spell out the relevant differences between these claims. Furthermore, because of the insufficient specification, in the philosophical literature, of the justificatory rationalist claim, I give my own characterization of it based on epistemological and ontological analogy that moral rationalist make with the rationality of mathematics. Finally, I give an overview of the relevant empirical studies that bear significance on the ideas of moral rationalism and base my arguments on these empirical findings. I conclude that empirical findings undermine most of the tenets of moral rationalism.

Table of Contents

LIST OF FIGURES

INTRODUCTION

In my thesis I will be dealing with the issue of the origin of our moral knowledge and the nature of moral judgment; more narrowly, I will examine the impact of the recent trends in moral psychology on contemporary debates in metaethics. In last couple of decades, as a consequence of discovering the importance of intuitive and affective processes in producing moral judgments, trends in moral psychology have moved from emphasizing the reasoning abilities in understanding moral knowledge to describing and investigating underlying intuitive (mostly unconscious) mechanisms and neural structures that play a role in social and moral cognition (Greene & Haidt 2002, Haidt 2007). This move in scientific investigation and discovery of importance of affective and intuitive processes has made a considerable impact on moral philosophers and inspired a significant interdisciplinary work between philosophers, psychologists and neuroscientists. Most notably, work in moral psychology has influenced and gave a new incentive in reconsidering philosophical conceptions of the relation between reason, moral knowledge, moral motivation, moral agency and its implications for the nature of moral knowledge and its modality (e.g. Greene 2008a, Nichols 2004, Prinz 2006). In this thesis I will align myself with those philosophers (e.g. Hume 1739/40, Nichols 2004, Prinz 2006) who claim that ambitions of moral rationalism are not feasible, and therefore I will argue that empirical data and theories based on them undermine some of the main tenets of moral rationalism.

Moral rationalism

Historically the most influential moral rationalist was Kant (1785, 1788). According to Kant, moral duties and accordingly moral judgments are based on the idea of practical reason, which means that moral requirements are requirements of practical rationality and that validity of moral norms can be justified by using our capacities for practical reasoning. This idea of basing

morality on norms of rationality is attractive since it enables one to explain and justify the objectivity of moral demands, and it also secures that every rational person can recognize the reason to act in accordance with moral requirements.

Three moral rationalist claims

In contemporary philosophy moral rationalism in some of its forms is defended by distinguished moral philosophers (e.g. Darwall 1982, Korsgaard 1986, 1996, 2008, Parfit forthcoming, Scanlon 1998, Smith 1994). However, not all of them would accept the same conceptions of the idea of basing morality on reason. Because of that, the basic rationalist idea that morality rests on practical reason can be unfolded into three kinds of claims about the connection between morality and rationality: conceptual, psychological[1] and justificatory claim.[2]

Conceptual rationalism is a descriptive claim about our concepts of moral requirements. According to the conceptual rationalist claim, it is conceptually true that moral requirements are requirements of practical reason (Korsgaard 1986, Smith 1994), or as Smith puts it, "our concept of a moral requirement is a concept of a reason for action'' (1994, p. 64). The claim is that, for example, when one judges that it is morally right to help starving children in Africa then one judges that she has a reason to help the starving children in Africa, and if one is not motivated to do that, then she is being irrational.[3] Hence, conceptual rationalism claims that it is a conceptual truth that all rational agents will, ceteris paribus[4], act in accordance with what she believes to be moral requirements.

[1] Nichols (2004, p. 67) introduced the distinction between conceptual and empirical rationalism; however, I will follow Joyce (2008) in calling the latter form of rationalist claim psychological rationalism.
[2] Joyce (2008) introduced the justificatory rationalism as a separate form of moral rationalism into the distinction made by Nichols (2004).
[3] Since in such a case the agent would fail to act in a way that she believes she has a reason to act.
[4] Ceteris paribus clause encompasses various forms of irrationality (weakness of the will, depressions, etc.).

Psychological rationalism is a descriptive claim about the capacities that produce moral judgments. According to the psychological rationalism, it is an empirical fact that moral judgments are a kind of rational judgments, i.e. they are derived from our rational capacities (Joyce 2008, Nichols 2004). Smith's analogy between morality and mathematics can be seen as an expression of this view; he says that "a convergence in mathematical practice lies behind our conviction that mathematical claims enjoy a privileged rational status", in the same way we can think that convergence in our moral practice would indicate that "moral judgments enjoy the same privileged rational status'' (Cited in Nichols 2004, p. 69).

When talking about rational capacities, philosophers usually do not specify what exactly they have in mind, but it can be supposed that under the term 'rational capacities' are implied cognitive processes that underlie our capacities for "reasoning, planning, manipulating information in working memory, controlling impulses, and "higher executive functions'' more generally.'' (Greene 2008a, p. 40)

However, the psychological rationalist claim that moral judgments are products of rational capacities cannot be construed as a simple claim about the causal antecedent of moral judgment. In reaching a moral judgment one has to respond to the available moral reason, where responding is not just causal. Otherwise it would have to be conceded that, for example, a dog trained to help avalanche victims is acting morally. In addition, it is not the case that psychological rationalism claims that moral judgment is *just* a causal output of cognitive processes. Cognitive processes underlying moral judgment might be simulated by a computer program; however, outputs of this program would not be a genuine moral judgment. That is why rationalists enrich the concept of responding to reasons with the *awareness* (Parfit forthcoming) of the normative (reason-giving) facts or with the recognition by self-reflection (Korsgaard 2008, p. 23) that certain consideration is a reason to perform some action. Parfit says that

> [w]e respond to decisive reasons when our *awareness* of the reason-giving facts leads us to believe, or want, or try to do what we have these reasons to believe, or want, or do. (…) To *fail* to respond to some reason, we must be aware of the facts that give us this reason. (Parfit forthcoming, p. 128, first emphasis added)

Korsgaard's (1984, 2008) view is more faithful to original Kantianism. She thinks that moral judgment is reached as a consequence of rational self-reflection. To judge whether some action is morally permissible, the agent has to adopt a maxim[5] and test whether it passes the categorical imperative test. For example, in thinking about whether to kill a person in order to have her revenge, an agent tries to universalize the maxim 'I will kill this person in order to have my revenge' in order to see whether it could become a universal policy according to which any rational agent can act (Korsgaard 2008, p. 218). Presumably, this attempt of universalizing will fail and the agent will reach a judgment that it is wrong to act on that maxim. So, in these views, moral judgments are products of conscious rational capacities processes, are contrasted with perception, sensation and emotion (ibid., p. 2).

Unlike conceptual and psychological rationalism, justificatory rationalism is a normative claim about the foundation of moral requirements. It claims that moral requirements are rational requirements, and that whoever does not act morally is susceptible to rational criticism. Justificatory rationalism needs to be distinguished from conceptual and psychological rationalism because it is not a conceptual or empirical claim about the sources of moral judgments but is a claim about the normative status of moral judgments and requirements.

However, justificatory rationalism is not adequately characterized in philosophical literature. For example, Joyce (2008) characterizes it as a view according to which "moral transgressions are rational transgressions; moral villains are irrational'' (p. 388). But this characterization is not helpful in distinguishing it from other rationalist claims, since even on conceptual or

[5] Korsgaard explains maxims as being descriptions of the action with the following structure ''to-this-act-for-the-sake-of-this-end''. (Korsgaard 2008, p. 218) In this view 'I will break promise that I made' is not a maxim, but 'I will break promise in order to gain some benefit' is a maxim that can be tested using categorical imperative.

psychological rationalist view it can be said that immoral agents exhibit some rational failings in their behavior.

We can give more concrete characterization of justificatory rationalism if we see what claims moral rationalists want to justify. One of the main motivations of moral rationalism is their desire to vindicate intuitions behind the claim that moral judgments express non-hypothetical[6], universal and impartial demands that have authority over all agents (Railton 2008). This aim is accomplished if we suppose that moral judgments express necessary truths that can be known a priori (Parfit forthcoming, p. 142, 678-9, Smith 1994, p. 192). So, I will construe justificatory rationalism as a defense of the thesis that moral judgments express a priori and necessary truths that can be grasped by using our rational capacities.

In the next three chapters I will examine the three rationalist claims. In chapter 1 I will present and discuss Nichols' (2002) argument against conceptual rationalism. In chapter 2 I will argue that empirical evidence and explanations of empirical data counts against the psychological rationalism. Finally, in chapter 3 I will argue that justificatory rationalism is not a plausible claim when we take into account the origins of moral knowledge and motivations that underlie our moral behavior.

[6] That is that they express categorical normative claims that apply to agents independently of their contingent goals or ends.

CHAPTER 1: CONCEPTUAL RATIONALISM

In examining the conceptual rationalism Smith's (1994) view comes to the fore, since he analyzes moral concepts in a way that will preserve the intuitions that people have about these concepts. His idea is that proper analysis of concepts will articulate all and only the platitudes that surround these concepts. For example, he says that

> in acquiring a concept C we come to acquire a whole set of inferential and judgmental dispositions connecting facts expressed in terms of the concept C with facts of other kinds. A statement of all of these various dispositions constitutes a set of platitudes surrounding C. (…) An analysis of a concept C in term of (…) C* is correct just in case knowledge of C* give us knowledge of all and only the platitudes surrounding C: that is, knowledge of all and only the inferential and judgmental dispositions of someone who has a mastery of the concept C. (Smith 1994, p. 37-38)

Here the most important point is that Smith considers conceptual rationalist claim and what he calls practicality requirement to be among the platitudes that surround our moral concepts (ibid., p. 39). According to the practicality requirement, it is "a conceptual truth that agents who make moral judgments are motivated accordingly, at least absent weakness of the will and the like.'' (Ibid., p. 66) The relation between the conceptual rationalist claim and the practicality requirement is one of entailment; according to Smith, the former entails the latter. To see this we may formulate conceptual rationalist claim in a following way: "if it is right for agents to Φ in circumstances C, then there is a reason for those agents to Φ in C.'' (ibid., p. 62) This means that when someone judges that some action is right, then that person judges that she has a reason to perform this action, but to judge that she has a reason to perform some action, according to Smith, is to judge that she would be so motivated if she were rational. Hence, if she judges that some action is right then, if rational, she would be motivated to act in accordance with it, and that is what practical requirement claims (ibid.)

From the practicality requirement it follows that a rational amoralist[7] is not possible. In real life, the closest that we come to rational amoralists are people with psychopathic disorder. They make moral claims, distinguish between right and wrong but then fail to act according to their moral judgments. Proponents of the practicality requirement claim that when psychopaths use moral terms, they use them in 'inverted-commas' sense (Smith 1994), so that when they express moral judgments and stay unmotivated to act according to them, it is claimed that they do not '*really* express moral judgments at all.'' (Ibid., p. 67) In order to secure the practical requirement, the inverted-commas response cannot be just an empirical fact about psychopaths, since that would not secure the practical requirement. Rather it must be a conceptual truth about moral judgment and it must follow from the practical requirement. This is because

> [c]onceptual rationalism is, after all, supposed to characterize our ordinary moral concepts and intuitions. Indeed, as Smith develops it, conceptual rationalism is supposed to be a systematized set of platitudes that characterize the folk concept of morality. (Nichols 2004, p. 73)

The connection between the practicality requirement and the conceptual rationalist claim is important because Nichols (2004) uses it to show that since practical requirement is false then by modus tollens so is the conceptual rationalist claim. In arguing for this conclusion he uses the folk conception of the psychopathic character in order to check whether it is a platitude surrounding our concept of moral judgment that psychopaths do not express real moral judgments. So, the idea is that if the claim that psychopaths do not really make moral judgments is part of our concept of moral judgment, then normal people that have competent mastery of moral terms should exhibit this platitude in applying their moral concepts. Nichols (2002) decided to test this claim. He conducted a study in which philosophically unsophisticated

[7] A rational amoralist is a person who delivers sincere moral judgments about what is right or wrong but stays unmotivated to act according to her judgments, without exhibiting any rational failings. A perfect amoral rationalist would be a person that embodies the devil.

undergraduates were asked to say whether a person from the presented story really understands moral claims. They were presented with the following probes:

> John is a psychopathic criminal. He is an adult of normal intelligence, but he has no emotional reaction to hurting other people. John has hurt and indeed killed other people when he has wanted to steal their money. He says that he knows that hurting others is wrong, but that he just doesn't care if he does things that are wrong. Does John really understand that hurting others is morally wrong?
>
> Bill is a mathematician. He is an adult of normal intelligence, but he has no emotional reaction to hurting other people. Nonetheless, Bill never hurts other people simply because he thinks that it is irrational to hurt others. He thinks that any rational person would be like him and not hurt other people. Does Bill really understand that hurting others is morally wrong?

Results of the study show that most of the subjects think that a psychopath did really understand that hurting others is morally wrong. On the other hand, most of the people answered that the mathematician did not really understand that hurting others is morally wrong. These results seem to show that it is not part of the concept of moral judgment that psychopaths do not express real moral judgments, which rather indicate that at least according to some people it is part of the concept of moral judgment that psychopaths really make moral judgments. This also indicates that practical requirement is not part of the concept of moral judgment and by implication the conceptual rationalist claim seems not to be one of the platitudes that surround our moral concepts.

1.1 Objections to Nichols' (2002) study

Joyce (2008) puts forward two important objections to Nichols' arguments against conceptual rationalism. The first objection is that Nichols' (2002) study does not show that the version of practical requirement that Smith (1994, 2008) endorses is not part of the concept of moral judgment. What Smith (1994) calls practical requirement is a version of what is known as judgment or motivation internalism. According to judgment (motivational) internalism, it is a conceptual truth that when one expresses a moral judgment, one is motivated or at least has some pro-attitude (inclination) to act according to that judgment. In the strong version of judgment internalism "if an agent judges that it is right for her to Φ in circumstances C then she is motivated to Φ in C''; thus on this version motivation is implied in moral judgment *simpliciter* (Smith 1994, p. 61).

However, the version that Smith endorses says: "if an agent judges that it is right for her to Φ in circumstances C, then either she is motivated to Φ in C or she is practically irrational''; in this version motivation is still internally connected to moral judgment, but there is still room for the connection to be impaired by "influences of the weakness of the will and other similar forms of practical unreason'' (ibid.). Since Nichols did not ask subjects in his study whether John, the psychopath, is irrational for not being motivated by his moral judgment, Joyce (2008, p. 382) claims that because of that Nichols' argument misses its target, and that it at most shows only that strong judgment internalism is not part of people's concept of moral judgment. This objection has a point and Nichols (2008) acknowledges that, but it only indicates that further studies need to be done in order to settle the issue concerning conceptual rationalism.

Joyce's other objection is that the type of study that Nichols (2002) conducted cannot illuminate the content of our concepts and thus, cannot say whether the product of a conceptual

analysis is part of our concept or not. Joyce (2008, p. 382) indicates that in Smith's account conceptual truths can be unobvious to ordinary speakers. According to Smith, to have a competence with a concept is to have certain inferential and judgmental dispositions with the concept, that is mastery of a concept requires us to have a knowledge-how, while ''knowledge of an analysis of a mastered concept requires us to have knowledge-that about our knowledge-how.'' (Smith 1994, p. 38) Joyce (2008) makes an analogy between a champion swimmer and a concept user and suggest that just as "it might be a bad way of figuring out how a champion swimmer swims by asking him to describe his swimming technique'' (p. 382), it can also be a bad idea to ask an ordinary speaker of a language to articulate the inferential and judgmental dispositions that underlie her use of some concept.

I think that this last objection does not hold. When we probe intuitions of certain ordinary speakers of some language, we do not expect them to articulate all inferential and judgmental dispositions that they have with the content of that intuition, but that is exactly what this kind of probing enables us to do (although in a limited way) to figure out what are the competencies underlying people's use of certain concepts. Nichols (2008) claims that Joyce's analogy with the champion swimmer is not adequate, since "we wouldn't expect folk platitudes to be the key source of information about the mechanics underlying a backstroke'', while on the other hand we can "expect the folk to recognize their own platitudes'', since the conceptual analysis in Smith's account is "supposed depend crucially on platitudes'' (p. 399). I agree with Nichols, and I think that the analogy with the champion swimmer misconstrues the problem. We do not have to ask people to describe their competence with the word, but rather we can infer what they mean by the word by asking them to use that word.

Also, I believe that the analogy can be construed in a better way; just as you discover swimmers swimming technique by observing them swimming, in the same way you discover

people's concept competences by observing how they use it. This can be conducted in conditions where the observer can isolate and manipulate variables that are important for the investigation. So, I would, in line with Nichols, conclude that probing folk's responses is a good way to outline the platitudes that surround certain concepts. However, even though there is no reason why we should not experimentally test whether proposed platitudes concerning moral concepts are really platitudes, still the discussion about whether conceptual rationalism holds, as conceived by Smith (1994), is inconclusive.

CHAPTER 2: EMPIRICAL MODELS OF MORAL JUDGMENTS

The basic conception of rationalistic moral psychology is that consciously reaching a moral judgment is a product of reasoning or reflecting about some moral issue. This idea was operationalized in psychology by Piaget (1965) and especially Kohlberg (1969, 1984). Kohlberg postulated six stages of moral development, which were grouped into three levels: preconventional, conventional and postconventional morality.

The idea was that at the preconventional level, at stage 1 children obey rules because of the fear of punishment; at stage 2 children start to recognize that there are different viewpoints from which actions could be assessed, and they start to relativize the rightness of different behaviors to individual's self-interest. At the third, conventional, level children become aware that there are expectations from their families and community members that they need to fulfill. The recognition of these expectations gets manifested in children's tendency to judge other people's motives as good or bad and in the desire to be well regarded by others. At stage 4 people start to adopt the perspective of the society as a whole; they become preoccupied with social stability, which is manifested in the emphasis they make in obeying the laws, social norms and respecting the authority. At the postconventional level people start to dissociate moral ideals form conventional norms. At stage 5 people justify obligatory actions by appealing to social contracts and democratic procedures through which people might determine what a society ought to be like. At the final, sixth stage people start to judge actions and assess society's moral status by invoking impartial and universal principles (e.g. principles of justice and rights) that transcend all social customs and ground them.

What is important about this supposed moral development is the idea that people supersede these stages, not as a function of maturational process, but as a consequence of people's ability to

reason about their moral views, and their capacity to make progressive improvements of their moral views thorough rational deliberation. Thus, according to the rationalist model (Figure 1), moral judgment is a causal effect of conscious reasoning. In that model affect or emotion can have a role in delivering a moral judgment, but it can only play a role in producing judgment as an input to moral reasoning, and it is supposed that it can never be a direct cause of the moral judgment.

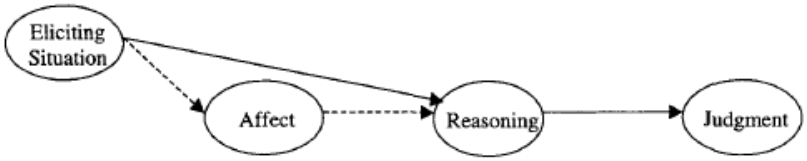

Figure 1: The rationalist model of moral judgment (Haidt 2001, p. 815)

However, it seems that there is strong evidence that points against this traditional rationalist model. In recent years three influential models of moral judgment have been developed (Greene 2008, Haidt 2001, Hauser et al. 2008). In the following sections I will present them and the data on which they are built and show how they undermine psychological rationalism.

2.1 Moral dumbfounding

Studies have shown (Nisbett & Wilson 1977) that when people explain their behavior, they often engage in constructing post hoc causal explanations because their actual causal processes are not accessible for conscious self-reflection. In constructing these explanations people turn to explanations of behavior that are supplied by their culture. These post hoc constructions of explanations of actions have been tested in studies in which experimenters hypnotize subjects to perform some actions and later ask why they were doing them. Usually people in those circumstances make up plausible reasons why they were doing the action in question; however,

these reasons are false, since they did not know why they were performing the actions they were hypnotized to perform.

Haidt (2001) argues that moral reasoning works in the same way. Haidt's general idea is that the same happens when one tries to explain her moral judgments. The idea is that after one reaches a moral judgment, then that person consults her culturally available moral theories, which give standards for evaluating the behavior that is judged. Since this process of reaching moral judgment is intuitive (inaccessible to consciousness), the most plausible explanation is that moral reasoning does not usually causally precede reaching a moral judgment.

Haidt and his colleges (cf. Haidt 2001) tested this hypothesis about moral reasoning. In their studies Haidt, Koller and Dias (cf. ibid.) conducted interviews with people in which they used a set of stories that would elicit strong affective reactions. They used stories like the following: a family eats its pet dog after the dog was killed by a car; a woman cuts up an old flag to create rags with which to clean her toilet; a man uses a chicken carcass for masturbation, and afterwards he cools and eats the carcass (cf. Haidt & Bjorklund 2008, p. 196). They noticed that people give their initial evaluations very quickly, but that they have some problems with giving supporting reasons for their judgments, which should not be a problem if judgments are products of moral reasoning. They discovered that when people's reasons are defeated, they continue to search harder for additional reasons and they rarely change their minds about the initial evaluation. Haidt calls this effect "moral dumbfounding'', because people tend to hold on to their moral judgments despite their inability to justify them.

Haidt, Bjorklund and Murphy (cf. ibid.) explicitly test the idea of moral dumbfounding. They gave subjects several tasks, for example, a behavioral task: a request to sip a glass of apple juice into which a sterilized dead cockroach had just been dipped. In each task the experimenter played

a ‘devil’s advocate’ and argued against anything the subject said. One of the vignettes that subjects received in this experiment was the following:

> Julie and Mark are brother and sister. They are travelling together in France on summer vacation from college. One night they are staying alone in a cabin near the beach. They decide that it would be interesting and fun if they tried making love. At the very least it would be a new experience for each of them. Julie was already taking birth control pills, but Mark uses a condom too, just to be safe. They both enjoy making love, but they decide not to do it again. They keep that night as a special secret, which makes them feel even closer to each other. What do you think about that? Was it OK for them to make love? (Haidt, 2001, p. 814)

After reading this vignette most of the subjects would answer that it is not Ok for Julie and Mark to have sex, but when the subjects were asked to justify their answer they would experience difficulty with finding a good reason for holding their judgment. Most of the subjects would try to argue that there is a possibility of inbreeding or causing psychological damage to the siblings. However, to this kind of responses the experimenter would answer that it is stipulated that there would be no harm to the siblings, and that they took all necessary precautions in order for Julie not to get pregnant. Similarly, for every other reason that subjects would give in order to justify the claim that incestual sex is wrong, the experimenter would give a counter argument why this answer is not a good response. In this task the idea of moral dumbfounding has been confirmed, since many subjects did not change their initial judgment even after they admitted that they could not explain the reasons for their moral judgments (Haidt & Bjorklund 2008, p. 198).

Further evidence for the primacy of intuition in reaching moral judgment comes from Wheatley and Haidt’s (2005) study. In that study they managed to manipulate people’s intuitions

which then affected their moral judgment. They hypnotized one group of subjects to feel disgust whenever they read the word 'take' and another group to feel disgust when they read the word 'often'. In study 1 subjects read six moral judgment stories, each of which included either the word 'take' or the word 'often'. In those studies subjects that felt a flash of disgust made more severe moral judgments than the ones that did not read the target word. More recently Schnall et al. (2008) demonstrated the reverse effect. Subjects whose cognitive concepts of cleanliness were activated and subjects "who physically cleansed themselves after experiencing disgust made less severe moral judgments relative to participants who were not exposed to cleanliness manipulations.'' (p. 1222) This study confirmed that when people make moral judgments they use intuitions, "even when these intuitions are incidental and irrelevant to the object or situation being judged.'' (Ibid.)

2.2 Haidt's social intuitionist model

In order to explain previous findings, Haidt (2001) proposed a social intuitionist model (SIM) of moral judgment. According to SIM, people typically reach moral judgments not as an product of moral reasoning, but as a result of "quick, automatic evaluations'', (ibid., p. 814) based on "emotions and affective intuitions'', so called "gut feelings'' (Greene & Haidt 2002, p. 517), which are typically construed as non-argumentative, unconscious processes, of which, only effects or final products (moral judgments) are available to conscious thought. In this respect Haidt (ibid.) indicates that Humean analogy between moral intuitions and aesthetic judgments is appropriate; "One sees or hears about a social event and one instantly feels approval or disapproval.'' (p. 818).

According to SIM, people reach moral judgments through aforementioned automatic affections (moral intuitions), and in this *production* moral reasoning is usually "an ex-post facto

process'' (ibid., p. 815) which people use to justify their pre-ordained conclusions (intuitive moral judgments). The metaphor with reasoning is that, after reaching a moral judgment, people act as lawyers who defend their intuitively reached judgments without caring whether this judgment might be false.

Haidt's SIM belongs to dual process theories of human cognition. According to these theories, there are two processing systems at work when a person makes a judgment: system 1 and system 2. System 1 includes processes that are associative, automatic, unconscious, fast, and context dependent. System 2 includes processes that are rule-based, controlled, conscious, serial and slow. System 1 constitutes the evaluative affective system that has primacy over System 2 in terms of phylogenetic and ontogenetic development. Haidt's proposal is that moral judgments are typically products of intuitive system 1, and that moral reasoning comes as a slow, effortful, domain general system 2. That is why Haidt defines moral intuition as quick, valenced (good-bad, like-dislike) and unconscious evaluation of social stimuli which results in moral judgment. On the other hand, moral reasoning is defined as conscious, intentional and effortful mental activity that consists of transforming given information about some object (ibid., p. 817-818). So, the idea is that when a person is confronted with a moral situation, e.g. as described in the incestual sex vignette, she will have a negative intuitive reaction (possibly a flash of disgust) that will produce a negative moral judgment. However, she will not be aware of the intuitive reaction and when asked for a justification of the moral judgment, she will engage in post-hoc reasoning in order to find some plausible argument of which the judgment will be a conclusion.

According to the social part of the SIM, people most often engage in conscious, reasoned deliberation when they are confronted with the demand to justify or explain their intuitively reached judgments. As a whole, this model consists of 6 links:

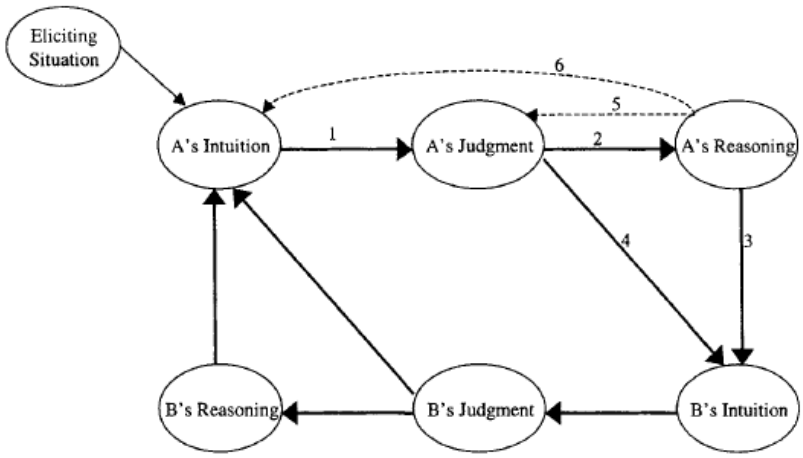

Figure 2: Social intuitionist model of moral judgment (Haidt 2001, p. 815)

1. Intuitive judgment link - moral judgment appears in consciousness automatically and effortlessly (result of the moral intuition).
2. Post-hoc reasoning link - after moral judgment is made, person engages in effortful process of searching for arguments that will support already-made judgment.
3. The reasoned persuasion link – moral reasoning is produced to verbally justify ones moral judgment to others.
4. The social persuasion link – members of a certain group (friends, allies, acquaintances), just by making moral judgments can affect others' moral judgments.
5. The reasoned judgment link – Haidt recognizes that at times people can override their own initial intuition. However, the model proposes that this kind of causal reasoning is rare, it can happen when "the initial intuition is weak and processing capacity is high.'' (Ibid., p. 819)
6. The private reflection link – by thinking things over or by role playing (putting oneself into the shoes of another) one can by experiencing sympathy, pain, etc. provoke a new intuition in oneself that will conflict with the initial intuition, which then needs to be

resolved, either by going with the stronger intuition or by letting reason decide on the basis of consciously applying some rule.

According to Haidt, unlike the rationalist model, which focuses on links 5 and 6, SIM, even though it also includes links 5 and 6, puts emphasis on links 1-4, while 5 and 6 may, but rarely do, contribute to the production of moral judgments.

2.3 Greene's model of moral judgment

Greene's model also belongs to dual process family of models; however, unlike Haidt's model, it is usually considered that it gives much greater role to reasoning processes in causal production of moral judgments. Greene constructed his model on evidence gathered by using functional magnetic resonance imaging machine (fMRI). In Greene et al. (2001, 2004) studies they administered fMRI scans to subjects that were making moral judgments on presented moral dilemmas. The dilemmas that they used are classical 'trolley problems'.

In the trolley dilemma there is a runaway trolley which is heading towards five people, who are working on a track and who will be surely killed if the trolley does not change its course. The only way to save these five persons is to divert the trolley, by flipping a switch, to a side track. However, on the side track there is one person who will, in the case of flipping a switch, surely be killed. The question is whether it is permissible to flip the switch. In the footbridge dilemma, there is again a runaway trolley that will kill five people if you do not do something. However, in this case you are standing on a footbridge, above the track, next to a large man, and the only way for you to save five persons is to push the large man in front of the trolley, thereby killing him. Here again the question is whether it is permissible to push the large man.

In answering these questions most people say that it is permissible to flip the switch in the first case, but that it is not permissible to push the overweight man on the track in the second case

(Hauser et al. 2008a). However, from the philosophical perspective, it is not obvious how to justify strong and different intuitions behind these dilemmas, since both cases involve sacrificing one person in order to save more, and the slight variations of the trolley problems prompts incompatible intuitions (see Hauser et al. 2008a). The question that Greene et al. (2001) wanted to investigate is why people have these strong intuitions about these dilemmas, and what psychological processes underlie the responses to these dilemmas. Their answer was that

> the crucial difference between the trolley dilemma and the footbridge dilemma lies in the latter's tendency to engage people's emotions in a way that the former does not. The thought of pushing someone to his death is, we propose, more emotionally salient than the thought of hitting a switch that will cause a trolley to produce similar consequences, and it is this emotional response that accounts for people's tendency to treat these cases differently. (Greene et al. 2001, p. 2106)

Greene et al. (2001) hypothesized that footbridge dilemmas prompt more emotional response because they put an agent into a situation that is more 'up close and personal' (pushing a man), while trolley dilemmas supposedly are more impersonal (flipping a switch), which does not make the emotional response salient; in fact they make reasoned response more salient. That is why they divided moral dilemmas that they presented to subjects into two groups: personal (e.g. footbridge) and impersonal (e.g. bystander) dilemmas.

Based on the idea that personal dilemmas provoke more emotional responses, Greene et al. (2001, 2004) predicted that thinking about personal moral dilemmas should produce more activity in brain areas underlying emotional processing, while thinking about impersonal dilemmas should produce more activity in brain areas underlying more cognitive processing. This prediction was confirmed; subjects who were delivering judgments on personal moral dilemmas had more brain activity in brain areas associated with emotion: the posterior cingulate cortex, the medial prefrontal cortex, and the amygdala; and subjects who were judging impersonal dilemmas had greater brain activity underlying cognitive processes: dorsolateral prefrontal cortex and inferior parietal lobe (cf. Greene 2008a, p. 44).

Greene et al. (2001) also predicted and confirmed that since emotional reactions are quick and automatic, someone who judges that personal moral violations are permissible (e.g. that it is ok to push an overweight man), will most likely have to override her initial emotional response against permitting this violation, which means that it would take relatively more time to answer 'yes' than 'no' to personal dilemmas. On the other hand, in the impersonal moral dilemmas (like the bystander dilemma) there is no expected default emotional reaction which needs to be overridden. So the prediction was that there will be no significant difference in response time, whether subjects answer 'yes' or 'no'.

Finally Greene et al. (2004) predicted that in difficult personal moral dilemmas (dilemmas that take more time to answer), where there is likely to be conflicting representations of behavioral responses there will be increased activity in more cognitive brain areas. To test this prediction, they presented to their subjects the *crying baby* dilemma. In wartime you find yourself with your fellow villagers hiding in the basement form enemy soldiers. Your baby starts to cry and you put a hand over her mouth. If you raise your hand enemy soldiers will find you and kill everyone in the basement. If you do not raise your hand soldiers will not find you, but the baby will surely die. The question is whether it is ok to kill your baby. To answer this question, people need relatively long time and there is no consensus between people's intuitions about this case (cf. Greene 2008a, p. 44).

On the other hand, people usually do not have trouble reaching a consensus about the *infanticide* dilemma. In this dilemma teenage girl must decide whether to kill her unwanted newborn. Most people give relatively quick answer that killing a newborn would be wrong (cf. ibid.). Greene and his colleagues hypothesized that these two cases of personal moral dilemma both include a quick negative emotional response to killing one's own baby; however in the

crying baby dilemma there is a conflicting more cognitive intuition based on consequentionalist[8] cost-benefit analysis, which favors killing your baby in order to save all other people. Since this conflict between intuitions is evident in response time, they predicted that the brain areas associated with response conflict and with cognitive processes will show increased activation. The prediction was confirmed, the anterior cingulate cortex, which is the brain area supposedly associated with response conflict (cf. ibid., p. 45), showed increased activation, and also the anterior dorsolateral prefrontal cortex and the inferior parietal lobes (associated with cognitive processes) showed greater activation compared to subjects that were answering the *infanticide* dilemma.

Hence, in Greene's account (see Figure 3) there are two processes underlying moral judgment, emotional and cognitive ones. They have different activation cues, they both play a causal role in producing moral judgment, and they can come into conflict.

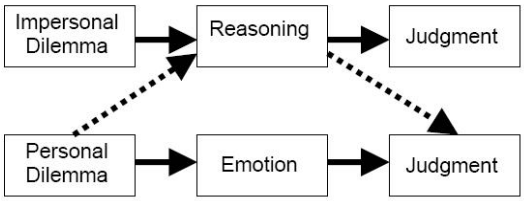

Figure 3: Greene's model (Nado et al. 2009, p. 9)[9]

2.4 ''Rawlsian'' model of moral judgment

Third influential model (see Figure 4) of moral judgment was proposed by different authors (e.g. Dwyer 2009, Hauser 2006) who based their ideas on an analogy with linguistic research

[8] Greene (2008a) connects consequentionalists judgments with impersonal moral dilemmas where there is no competing negative emotional response (personal dilemmas), and claims that in impersonal dilemmas we reach a moral judgment by using cost-benefit analyzes of the outcomes of various possible acts.

[9] References to this article are from: http://www.rci.rutgers.edu/~stich/Publications/Papers/Moral%20Judgment%20-%20FINAL%20DRAFT%20-%20web.pdf

program developed in the Chomskian tradition. Inspiration for this model comes from remarks that John Rawls made in his book *A theory of Justice*, about the possible analogy between our linguistic competence and our moral competence. Basic idea behind this research program is that just like there is a language faculty that is described by principles of Universal Grammar, there is a moral faculty that is described by principles of *moral grammar*. Principles of moral grammar are thought to be innate and universal (species-specific); they specify parameters or constraints for adopting particular moral rules that are present in the environment in which the child grows up. Similarly to the language faculty, proponents of the linguistic analogy contend that moral faculty "operates unconsciously, quickly, and automatically'' (Nado et al. 2009, p. 10, see Hauser 2006, Hauser et al. 2008a), and that analogous to linguistic study, in which linguists use grammaticality judgments in order to discover principles underlying language competence, proponents of the linguistic analogy claim that by probing agent's moral judgments one may "uncover some of the principles underlying our judgments of what is morally right and wrong.'' (Hauser et al. 2008a, p. 118) In this respect Rawlsian model is intuitionist; however, unlike Haidt's model, proponents of the linguistic analogy consider these moral intuitions to be independent of any emotional processes (Dwyer 2009, Hauser 2006).

In order to develop more thoroughly the linguistic analogy, proponents of the Rawlsian model have put emphasis on the necessity for providing a description of computations underlying moral intuitions and judgments (Hauser et al. 2008a, Mikhail 2008). The basic idea of this model is that there is an appraisal system which, after perceiving an action, computes its causal-intentional structure and accordingly delivers a moral verdict about the permissibility of the action under evaluation. This process of action analysis "constitutes the heart of the moral faculty'' (Hauser et al. 2008a, p. 117), and it operates on the basis of tacit principles which constitute the moral grammar. It is hypothesized that this innate moral module works

independently of deliberate reasoning and emotional processing, and it generates moral judgment which then may (but it does not have to) be followed by emotional reaction or deliberate reasoning.

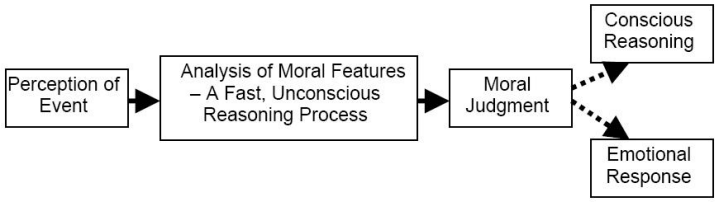

Figure 4: ''Rawlsian'' model (Nado et al. 2009, p. 10)

Most evidence that favors ideas proposed by proponents of the linguistic analogy comes from studying trolley problems.[10] Testing people with trolley dilemmas, proponents of the Rawlsian model (cf. ibid.) managed to show that when delivering moral judgments people are sensitive to features of the situation that cannot simply be explained by invoking different emotional responses to the situation.

The basic contention of the proponents of the linguistic analogy is that when people judge permissibility of particular actions then the issue is not whether the judgments spring from emotional or rational processes but which are the principles of moral grammar that shape these intuitions. It is not an easy task to formulate or even decide which principles constitute moral grammar[11]; however, there were some proposals for the principles that could play a role in moral grammar; for example, some version of the doctrine of double effect[12] can be seen as being a part of moral grammar. Using these sorts of principles proponents of the linguistic analogy try to

[10] Some of the proponents of the Rawlsian model use 'poverty of the stimulus' argument to argue that moral knowledge is innate. However, I am not going to enter into the discussion whether this argument is plausible, since it is not important form my discussion whether moral knowledge is innate or not.

[11] For example Dwyer (2008) explicitly states that "the form and content of the principles that I claim characterize the moral faculty remain a mystery.'' (p. 414)

[12] Cushman et al. (2006, p. 183) define the principle in the following way: harm intended as the means to a goal is morally worse than equivalent harm foreseen as the side effect of a goal.

explain people's intuitions in trolley dilemmas by focusing on the computations of causal-intentional structure of the contemplated action rather than on (like in Greene's account) emotional reactions that these dilemmas might or might not elicit. For example, Mikhail (2008) postulates that moral faculty in producing a moral judgment involves four different kinds of operations that compute the relevant features of an action: first (1) it identifies the various action descriptions and places them in an appropriate temporal and causal order; then (2) it applies certain moral principles to these action descriptions in order "to generate representations of good and bad effects''; next (3) it computes "the intentional structure of the relevant acts and omissions''; and (4) it derives more fine-grained representations of morally salient acts[13] (p. 87).

While Greene explains people's reactions in footbridge and bystander cases by invoking the personal-impersonal distinction, Rawlsian model predicts that in footbridge cases a person will represent the causal-intentional structure of the action as pushing that is impermissible because the man is wrongly used as a means without his consent, while flipping the switch in the bystander cases will be represented as permissible, since the man dies as a side-effect of the intended action of saving the five people.

In order to test their predictions and "to probe the nature of our appraisal system'' (Hauser et al. 2008a, p. 127) proponents of the Rawlsian model devised variations on the footbridge scenario. There were two important variations; in one scenario Ned is walking near the trolley tracks when he notices that there is a runaway trolley that is heading towards five people that were on the tracks. Fortunately Ned is standing next to a switch which can divert the train to go to a side-track; however, the side-track is a loop on which an overweight man is standing, and if

[13] Here Mikhail has in mind cases like the footbridge, in which he postulates that a person "must infer that the agent must *touch* and *move* the man in order to throw him onto the track in the path of the train, and the man would not *consent* to being touched and moved in this manner because of his interest in self-preservation" (Mikhail 2008, p. 91, f. 4).

it were not for the overweight man, the trolley would just loop back to the main track and kill the five people. However, if Ned flips the switch and diverts the trolley, the trolley would stop, but the overweight man would be killed, and he would be used as a means to save the five people. In the second scenario Oscar finds himself in a similar situation, the only difference is that on the side-track there is a heavy weight that could stop the trolley and in front of it there is a man who is not large enough to stop the trolley. If Oscar flips the switch, the trolley would be stopped, but the man would die as a side-effect of Oscar's trying to save the five people. The question is whether it is permissible for Ned and Oscar to turn the switch.

Hauser et al. (2008a) report the results from their cross-cultural internet study[14]: in the Ned scenario 55%, while in the Oscar scenario 72% of subjects replied that it is permissible to flip the switch. This difference is significant and it is not expected when using Greene's personal-impersonal distinction. Since both of these cases should fall under the impersonal category, there should be no emotional reaction that needs to be countervailed; therefore, according to Greene's account, there should not be such a significant difference in people's reactions in these two cases. In the Rawlsian model the difference is expected, since it predicts that people's appraisal system analyzes the causal-intentional structure of actions and it notes the difference between killing a man as a necessary means to save the five and killing a man as a by-product of a good consequence (saving the five people).

This study shows that there is a significant amount of cognitive processing going on, that makes it possible to make fine-grained distinctions before a person delivers a moral judgment. Moreover, this process is mostly unconscious and intuitive, since when subjects were asked to justify their answers to these two moral dilemmas, 87% could not provide a sufficient justification for making the distinctions in these two cases (ibid., p. 133). If the judgments were

[14] They gathered their data from the Moral Sense Test: http://moral.whj.harvard.edu/ .

reached by conscious reasoning, then these people would be able to provide the principle based on which they were reasoning.

2.4.1 Weak and strong linguistic analogy

Proponents of the Rawlsian model introduced a compelling argument[15] that before reaching a moral judgment people often go through complex, unconscious cognitive processes that analyze the situation to which a subject responds to. It seems that in order to have a complete account of moral cognition one would have to take into account the computations that underlie the extraction of information from the environment that are then used for reaching a moral judgment. This is what Hauser et al. (2008a) call the weak linguistic analogy. According to the weak linguistic analogy "[m]inimally, each of the other models must recognize an appraisal system that computes the causal-intentional structure of an agent's action'' (ibid., p. 117) which then can lead "to an emotion or process of deliberate reasoning.'' (Ibid., p. 121) This requirement can be accommodated to fit Haidt's and Greene's account, since this kind of appraisal system does not seem to be specific to the moral domain. The system that computes the causal-intentional structure of actions and assigns them descriptions does not have to be specifically associated with the system that applies moral principles to action descriptions; the system can "assign descriptions to actions toward which we have no moral reaction'' (Mallon 2008, p. 146). Also, it seems more plausible to think of this system as being more generally connected with systems that underlie our mind-reading abilities (theory of mind[16]).

However, in the strong linguistic analogy appraisal system "represents our moral competence and is responsible for the judgment'', which "then either triggers or doesn't trigger emotions and

[15] See section 2.4, p. 24-27.

[16] Theory of mind is an ability to ascribe and infer mental states (desires, beliefs, intentions, etc.) of oneself or other people, and to use it to explain and predict people's behavior.

deliberate reasoning.'' (Hauser et al. 2008a, p. 121) So, here emotions can only influence the performance of the moral faculty but do not function as part of the moral competence. In the strong linguistic analogy, like language faculty, moral faculty is conceived as being a specialized, dedicated and encapsulated system that works on unconscious principles which put a constraint on the range of learnable moral systems (ibid., p. 120, 139), and it is expected to exhibit "selective breakdowns due to damage to particular areas of the brain'' (ibid., p. 139). Mallon (2008) interprets the linguistic analogy as claiming that specialized moral faculty is a functionally distinct mental subsystem that involves: (1) proper functioning in the domain of morality; (2) computations that are performed on the limited sort of information (encapsulation) that are governed by principles which are opaque to conscious thought; and it is speculated that it might have (3) a particular brain location, i.e. that is physiologically discrete (the moral organ) (p. 146). Therefore, according to the strong linguistic analogy, Rawlsian model is incompatible with Haidt's and Greene's models.

Even though Rawlsian model represents an important research program, I do not believe that strong linguistic analogy is plausible. Because of the lack of space I will present only the most significant studies that count against the strong linguistic analogy. Koenigs et al. (2007) presented to patients with ventromedial prefrontal cortex (brain area connected with emotions) damage a large class of moral dilemmas. Among the presented dilemmas there were the standard trolley problems. What is important here is that when presented with 'personal' scenarios (such as the footbridge dilemma), these patients give significantly more consequentialist answers than the control subjects. That is, in cases like the footbridge scenario patients with impaired emotional processing make choices that lead to more aggregate welfare of people, choosing to sacrifice one in order to save more. This gives credibility to the claim that at least in some cases emotions seem to be causally necessary for reaching a moral judgment, since emotional

impairment leads people to exhibit abnormal pattern of moral judgment (compared to the controlled group). Greene (2008b, p. 138-139) mentions a study that also showed that patients with frontotemporal dementia[17], when presented with 'personal' moral dilemmas, tend to give consequentialist answers significantly more often than the controls, which is expected, since they have impaired emotional processes. These studies go against the idea that emotions are only consequences of moral judgments, since impairments in the brain areas connected with emotions imply impairments in moral judgment.

Complex processes that are involved in production of moral judgment indicate that there is no specialized moral faculty which would be computationally and physiologically discrete. Neuroimaging studies have shown that brain areas underlying emotional and cognitive processes implicated in moral judgment are also implicated in other activities such as non-moral social cognition, theory of mind, decision-making (Greene et al. 2001, Greene & Haidt 2002, Greene et al. 2004). Also, the fact that computations of the casual and intentional structure of actions are more naturally connected to mind-reading abilities, which then inform influence our moral judgments, and the fact that moral intuitions can have a reversed influence on our theory of mind judgments (see e.g. Knobe 2006) indicates that there is no computationally discrete or encapsulated moral faculty.

2.5 Implications of the empirical data for the psychological rationalism

Psychological rationalism claims that moral judgments are products of rational capacities, where this rational production is conceived as a process through which one reasons oneself into reaching a moral judgment. Psychological rationalism can be illustrated in two ways. For example, Korsgaard (2008) following Kant believes that reaching a moral judgment includes self-

[17] This impairment is also connected to impairments in emotional processing.

reflectively testing a maxim using the categorical imperative.[18] Testing a maxim through categorical imperative presumably presupposes activation of higher cognitive abilities, and conscious reasoning since it is postulated that the testing of maxims works through self-reflection (Korsgaard 1996, 2008), and self-reflection is a paradigmatic cognitive activity that requires conscious awareness.

On the other hand, according to Parfit (forthcoming, p. 156) moral judgments are judgments about what we could rationally want, where rationality consists in being aware and responding to reasons that there are for being moral (ibid., p. 128). Parfit's claim about *awareness of* and *responding to* reasons in connection to moral judgments can be construed in two ways: it is possible that one's awareness that there is a reason to flip the switch in Ned's[19] case causes one to produce a moral judgment that it is permissible to flip the switch, where this responsiveness to reason and reaching a moral judgment is intuitive (does not involve conscious reasoning); or it can be read that awareness and responsiveness to reason include conscious application of rational capacities. I would argue for the second option, since in Parfit's (forthcoming, p. 679) view we do not causally respond to reasons[20], and because of that, awareness of reasons, which leads to moral judgments, cannot be causal. For example, when someone in Ned's case delivers a moral judgment that it is permissible to flip the switch, she first has to recognize that there is a feature of the situation that gives her a reason to flip the switch (e.g. the fact that flipping will save more lives), and the awareness of that reason-giving fact leads her to judge that the act is permissible. Since that reason-giving fact is not causally available (ibid.), it cannot be the case that responding to it and the awareness of it is causal. That is why it is implausible to suppose that Parfit's awareness and responding to reasons can be construed as an intuition in Haidt's or Rawlsian

[18] See page 4 and footnote 5.
[19] See page 25-26.
[20] See section 3.3.

model, because those intuitions are causal consequences of stimuli form environment that satisfies some input conditions. Since Parfit does not postulate any special rational faculty that awareness of and responding to reasons might consist in (ibid.), it is more plausible to suppose that in his view, we reach moral judgments through rational deliberation, construed as a kind of domain-general reasoning ability.

In addition to being committed to a view of moral judgment according to which moral judgments are products of domain-general reasoning abilities, Korsgaard's and Parfit's view of moral judgments has as a consequence the claim that people who deliver moral judgments will have sufficient justification for them. This follows from the fact that according to their views, people if rational, will reach a judgment as a consequence of reasoning about the moral situation.

However, psychological rationalism to which moral rationalists are committed seems undermined by empirical data. I will summarize the main data from this chapter that counts against psychological rationalism. Moral dumbfounding and manipulation with people's intuitions indicate that conscious reasoning abilities are not primary sources of moral judgments, and that more often moral reasoning is motivated by a need to rationalize preordained moral judgments. In this process of rationalization people extract justifications from culturally supllied moral theories, which in addition make them biased to justify their own prejudices. Also, people rarely know the reasons in accordance with which they make moral judgments, which shows that they do not consciously reason themselves into making those judgments. Furthermore, complex computations that underlie production of moral judgments indicate that normal moral cognition depends on fine-grained distinctions that can hardly be captured by using domain-general reasoning capacities. Moreover, moral judgments are often produced by emotional processes, rather than cognitive ones, and emotions seem to be necessary ingredients in producing normal moral judgments, since impairments in emotional processing cause defects in moral judgment.

At this point, one might argue that studies, such as ones, made by Koenigs et al. (2007) with patients with emotional deficits (impaired ventromedial prefrontal cortex) show that rational capacities are after all sufficient for delivering moral judgments. However, we can say that these patients are making moral judgments, but these judgments must be considered defective, since they decline from exhibiting normal moral judgments and tend to give significantly more often consequentialist answers. Furthermore, one might claim that this is not a problem because these judgments are not affected by emotions and that these judgments are now more rational than they were before the injury.[21] However, this answer is not plausible, since there are some indications that people with the same emotional deficits do not become more 'rational' and less emotional. Koenigs and Tranel (2007) showed that ventromedial prefrontal cortex patients in ultimatum games[22] make more emotional and economically irrational choices compared to control groups, which indicates that people with these impairments exhibit defective moral judgments which cannot be ascribed to enhanced rational abilities.

Moreover, evidence from people with emotional impairments often exhibit sever deficits in moral functioning. More concrete test cases for this claim are based on studies made on psychopaths and people with 'acquired sociopathy'.

2.5.1 'Acquired sociopathy' and psychopaths

'Acquired sociopathy' is a term that was introduced by Damasio (1994) for adult patients who suffer from impairment in ventromedial prefrontal cortex area of the brain. Patients with these impairments usually show no reduction in their reasoning abilities; they retain knowledge

[21] See (e.g. Greene 2008a).

[22] Ultimatum games usually consist of two players who are given the opportunity to split the given money. One player makes an offer and if the other player (responder) accepts the offer then they both can keep the proposed split; if the responder refuses the offer then nobody gets anything. Usually when responders get an unfair offer (e.g. proposer get 10 dollars to share and offers 3 or less) they decline the offer because of the emotional reaction (anger) towards the unfair offer, which is economically irrational since to get some things better than to get nothing (Koenigs & Tranel 2007).

of moral rules and social conventions, and they can reason about hypothetical moral situations. However, these patients show various impairments connected to deficits in emotional processing. They show impairments in recognizing emotional expressions connected with anger and embarrassment (Blair 2003). They exhibit deficits in social behavior; they have problems with tolerating frustration, they tend to overreact and show reactive aggression to seemingly unimportant provocations (Blair 2003). When presented with emotionally charged pictures (e.g. of people dying, mutilation, social disaster), they "fail to show autonomic responses'' (Blair & Cipolotti 2000, p. 1123), such as the skin conductance response, which in non-patients indicates emotional arousal and responsiveness to morally salient situations. These patients also show deficits in decision-making as a consequence of emotional deficits (Damasio 1994), which means that even their capacity for practical reasoning gets defective as a consequence of emotional impairments.

In contrast to 'acquired sociopathy', psychopathy is a developmental disorder. Psychopaths are characterized as people who have good intelligence (Kiehl 2008), do not have delusions and do not think irrationally (Haidt 2001, p. 824). However, it seems that their reasoning capacities are being dissociated from their moral emotions (ibid.). They know the rules of social behavior and they understand the consequences of their actions, but they just do not care about acting morally and for moral reasons; and more importantly, they lack feelings of remorse, sympathy, empathy, shame, embarrassment. Characteristic for psychopaths is the callous, "goal-directed instrumental aggression and antisocial behaviour'' (Blair 2003, p. 7). Like people with 'acquired sociopathy', psychopaths do not react to pictures or images that cause distress in normal observers, also "they experience pain less intensely than normal subjects'' (Prinz 2006, p. 32) and beside impairments in recognition of emotions in facial expressions, they have trouble in reacting to emotionally charged words and sounds (Kiehl 2008).

Empirical evidence suggests that psychopaths also have impaired moral judgment as a consequence of their emotional impairments. Psychopathy is a developmental disorder connected to deficits in functioning of the paralimbic area of the brain (ibid.) and especially with early dysfunction of amygdala (Blair et al. 2001), which causes impairments in recognition of fearful and sad expressions. Blair's (cf. 2003) suggestion is that impairments in recognition of emotional expressions make children with psychopathic tendencies unable to experience fear and sadness as aversive unconditioned stimuli. "As a consequence of this, the individual does not learn to avoid committing behaviours that cause harm to others and will commit them if, by doing them, he receives reward.'' (Blair 2003, p. 6) This impairment in recognizing and experiencing certain emotions "leads [children with psychopathic tendencies] to a failure in socialization.'' (Ibid., p. 7)

Furthermore, Blair (cf. Nichols 2004) presented evidence that psychopaths fail to make moral-conventional distinctions. In the moral-conventional task subjects are presented with vignettes containing moral and conventional transgressions. Moral transgressions include acts like pulling one's hair, and conventional transgressions include chewing gum at school. The difference between these two transgressions consist in permissibility (moral violations are less permissible), seriousness (conventional violations are more serious), authority (moral rules are not authority dependent[23]) and the nature of justification.[24] Even three year olds are able to make moral/conventional distinction.[25] However, psychopaths are not. Rather they tend to treat all violations at the same level. For example, when they are asked to explain why they find certain moral violations prohibited they make significantly less reference to people's welfare even in the

[23] For example hitting someone (moral violation) would be judged as wrong, even if the parents say that it is ok to hit someone. In the same time, chewing gum in school will be judged as ok, if the teacher says that it is permissible.
[24] In justifying moral rules people will most often invoke concern for the well-being others, while in justification of conventional rule one will often appeal to social order or avoidance of punishment.
[25] Also people with autism and Down syndrome make the distinction.

cases where the harm to victims was clear. Since the biggest difference between normal people and psychopaths is thought to consist in psychopath's impaired emotional processing, it is plausible to conclude that the failure in psychopath's reaching a genuine moral judgment is the failure in her emotional capacities (Nichols 2004, Prinz 2006).

So, in addition to evidence from the previous sections,[26] the evidence presented here suggests that antisocial behavior is a consequence of emotional deficits and that proper emotional functioning is necessary for normal moral development and understanding. Hence, it seems that rational capacities are not sufficient for explaining moral understanding and judgment, and therefore, that psychological rationalism is false.

[26] See section 2.5.

CHAPTER 3: JUSTIFICATORY RATIONALISM

There are some pervasive features of moral domain and practice that are in need of explaining. For example, Peter Railton (2008, p. 38) mentions the following: non-hypothetical character of moral judgments – moral judgments are categorical requirements whose authority does not depend on agents contingent ends, desires, beliefs, etc. The nonrelativistic character of moral assessment, that is, the objectivity of moral judgments – " 'When A says that Φ-ing is right, and B says that Φ-ing is not right, then at most one of A and B is correct' '' (Smith 1994, p. 39). Further feature is the inability to settle moral disputes by using empirical means.

Justificatory rationalism purports to explain these features by attaching the moral domain to domain of a priori and necessary truths which can be discovered and justified using only our reasoning abilities, and in that way secure categoricity, and objectivity of moral requirements. Hence, justificatory rationalism claims that there is a "rational foundation for the content of our moral judgments.'' (Nichols 2008, p. 405)

3.1 The problem of contingent justification

In a trivial way, even on Haidt's account moral judgments can be rationally justified; for example, by constructing a post-hoc rationalization of one's intuitive moral judgment, or by engaging in a private or social reasoning process. Indeed, this is what philosophers usually do; they take intuitive moral beliefs that people have and try to systematically justify them by building a moral theory.

Haidt and Bjorklund (2008) especially emphasize the importance of social context for moral reasoning. They say "people are very bad at questioning their own assumptions and judgments, but in moral discourse other people do this for us.'' (p. 193) In that respect Scanlon's (1998) moral contractualism is congenial since it construes justification of moral judgments as a process

in which people who are motivated "to find principles for the general regulation of behavior'' try to find principles "that others, similarly motivated, could not reasonably reject.'' (p. 4) So, the idea is that when I judge that some action is wrong, then that "action would be one that I could not justify to others on grounds I could expect them to accept.'' (Ibid.) However, the empirical problem with this suggestion (that the moral rationalist wants to avoid) is that people who are not already motivated to justify their actions to other people would seem to have no reason to act morally. Moreover, it is an open question which justifications people can reasonably accept. Human 'ultrasociality' makes people sensitive and naturally predisposed to be responsive to other people's feeling and attitudes, which disposes people to reach consensus on moral issues (at least when they are part of the same community). From the evolutionary perspective this is explained by benefits (increased fitness) that the individuals will get from cooperation with other members of their community (Haidt 2007). Also, it is hypothesized that in order to benefit from cooperation one must track the reputations of others and manipulate with others in order to enhance one's own reputation. One of the most ubiquitous ways to track reputations is by gossiping;

> [i]n gossip people work out shared understandings of right and wrong, they strengthen relationships, and they engage in subtle or not-so-subtle acts of social influence to bolster the reputations of themselves and their friends. (Haidt &Bjorklund 2008, p. 190)

From this perspective it is clear that justifications and reasonings that people will offer, and are inclined to settle on, will not reflect the objective moral reality, but will rather reflect the motivations and cultural embeddings in which a person functions. I will call this kind of justification of moral conduct contingent justification, since what it justifies is a matter contingent on people's evolution, history and aims.

To avoid this possibility of *contingent* rational justification, moral rationalist invokes idealizations of the conditions of full rationality (Smith 1994), or postulating the non-natural

domain of irreducibly normative facts that we comprehend by using rationality (Parfit forthcoming). In the following sections I will consider these two views.

3.2 Justification under conditions of full rationality

On Smith's account[27] when someone judges that it is right to Φ then one judges that there is a normative reason to Φ. To judge that there is a normative reason to Φ is to believe that if one were a fully rational agent, one would desire to Φ. Full rationality includes: that the agent (1) has no false beliefs, (2) has all relevant true beliefs, and (3) deliberates correctly (Smith 1994, p. 156). The most important part of full rationality is correct deliberation where this includes a process of systematic justification of our desires (ibid., p. 156). This process is construed as one in which a person is trying to reach a reflective equilibrium, where this includes through revision accomplishing unity and coherence between one's desires and evaluative beliefs. Smith's thesis is that if we were all fully rational then our moral judgments and theories would converge. However, Smith bases his thesis as an inference to the best explanation from the empirical evidence:

> empirical fact that moral argument tends to elicit the agreement of our fellows gives us reason to believe that there will be a convergence in our desires under conditions of full rationality. For the best explanation of that tendency is our convergence upon a set of extremely unobvious *a priori* moral truths.[28] (Smith 1994., p. 187)

Smith gives examples of convergence in moral opinions that were *inter alia* obtained "via a process of moral argument'': "debates over slavery, worker's rights, women's rights, democracy and the like.'' (Ibid., p. 188)

I believe that Smith is wrong in assuming that tendency in convergence is best explained by the rational grounding of moral judgments. If there is a tendency towards convergence in moral opinion then there are other more plausible explanations. As I tried to show in the previous

[27] See chapter 2 of this thesis.

[28] I assume that the last sentence in quote means that what best explains convergence is the rational status of morality (see Nichols 2008, p. 404-405).

chapter, intuitive and emotional processes are necessary in moral development and moral understanding, while rational capacities do not seem to be sufficient. In that light it would be more plausible to explain convergence (past or future) as a consequence of similarity in intuitive and emotional repertoires that we as members of a single species have (Nichols 2008, p. 405). Smith could argue that even in that case morality could have rational foundation, because we use moral arguments and we are often responsive to them. This is true, but it will not secure justificatory rationalist claim; since if our moral intuitions or emotions were different then what we would argue for and what convergence we might reach could possibly be radically different, which would make claims about apriority and necessity of moral truths empty. We can just imagine how our moral world would look like if everyone had only consequentialist intuitions and were emotionally detached so that every moral situation was of the bystander dilemma kind. From my perspective such a world would be morally defective.[29]

Also, convergence might plausibly be explained by invoking evolutionary benefits of cooperation and its extension through processes, such as globalization, through which people get more and more homogenous in their social, moral and cultural patterns of behavior and opinion. Cause of this convergence would be then plausibly explained by cultural and economic transmission, innate biases, social pressures and not simply by rational argumentation. Therefore I contend that Smith's contention that convergence in moral opinion is best explained by invoking justificatory rationalist claim is not very plausible.

3.3 Justificatory normative realism

Similarly like Scanlon, Parfit tries to justify moral judgments by appealing to principles "whose universal acceptance everyone could rationally will.'' (Parfit forthcoming, p. 311) Where

[29] However, it would not seem to be rationally defective.

'what everyone could rationally will' means what 'everyone would have sufficient reason to choose'. What everyone would have sufficient reason to choose, according to Parfit depends on mind-independent, irreducibly normative facts, whose paradigmatic form is *being a reason for*. So, when answering a question whether some act is right or wrong we are looking for what we would have the most or sufficient reason to do. The only way to respond to these reason-giving facts is by using our rational capacities, since there is no causal connection with those non-natural facts (ibid., p. 676-678). By introducing these normative facts about our reasons, the contingency of justification is avoided, since it is supposed that normative facts correspond to necessary truths.

It is not clear how can moral judgments (and other normative judgments) be justified by facts to which there is no causal link, how can we even know about such facts? Parfit is aware of this epistemological problem. He considers the *Massive Coincidence Argument*: since there is no natural connection between our beliefs and normative facts, it is not clear how we can have so many true beliefs about normative facts. If there is no available explanation of this correlation, then we can only assume that the existence of this correlation would be a massive coincidence. The occurrences of such a correlation is highly unlikely, therefore we have no reason to assume the existence of normative facts (ibid., p. 678).

In an attempt to show that the latter argument is not successful Parfit makes the following analogy: we can make computers that can reliably produce true answers to mathematical questions, and they can do that without having any causal relation with numbers and their properties. Similarly, Parfit claims "[a]s the facts about computers show, we might be able to respond to such reasons without being causally affected by the normative properties of the reason-giving facts.'' (Ibid., p. 679) He even speculates that this rational ability to non-causally respond to reasons was selected for by natural selection.

One way in which this analogy could make sense is the following: computers respond to their inputs which then causally, according to some program, produce outputs. They are reliable because we built them. In case of morality we can suppose that the Rawlsian model (on the weaker linguistic analogy) developed as an adaptation. Then, we have inputs that get activated by stimuli from the environment which causally produces some moral judgment. The mechanism is reliable because it is an adaptation and functions normally. However, this is not enough, since it does not guarantee responsiveness to normative facts, because which principles actually govern the mechanism is contingent on our evolutionary history.[30] So, the output of this mechanism would have to be justified in some way; presumably by using domain general reasoning abilities.[31] However, in that case, my claim is again that this move cannot secure the a priori and necessary status of moral judgments, because we can plausibly suppose that this justification would be contingent on the nature of us as people.

For example, Greene (2008a) accounts for the difference between personal and impersonal forms of harms, as stated in trolley problems[32], in evolutionary terms. Crudely, the explanation is that given the fact that personal violence is evolutionary more ancient, and that affective system predates capacities for more abstract reasoning, then it is likely that humans evolved alarmlike emotional responses "to certain basic forms of interpersonal violence, where these responses evolved as a means of regulating the behavior of creatures (…) whose survival depends on cooperation and individual restraint." (p. 43) In cases where the violence is more impersonal, this alarmlike emotional response will not be triggered, which would allow people "to respond in a more 'cognitive' way, perhaps employing a cost-benefit analysis.'' (Ibid.) With this evolutionary

[30] We can again imagine that all humans have dominantly consequentialist's intuitions in cases like footbridge dilemma.
[31] Parfit claims that the ability to respond to reasons does not involve some 'mysterious' or 'quasi-sensory' intuition (forthcoming, p. 679), so I presume that rational abilities refer to our normal rational capacities.
[32] See section 3.3.

background Greene explains the difference in responses that people have to footbridge and bystander dilemmas. Now, most people find pushing a man of the footbridge to be wrong, while they do not find wrong flipping a switch in the bystander dilemma. This difference can be justified by invoking a moral principle that says it is wrong to kill someone as means of benefitting someone else, but that it is not wrong to kill someone as a side-effect of benefiting someone else (Parfit forthcoming, p. 329). However, we can imagine that our intuitions were different; for example we could have strong intuitions that killings in both dilemmas are equally wrong; then, presumably we could find some principle that would justify those intuitions. So, if we realize the contingency of our moral judgments on our nature then it seems less plausible to consider their justification to indicate the a priori and necessary status of moral claims. I am not claiming that our intuitions are irrational; on the contrary they might be very rational and advantageous for individual's fitness. However, this will not indicate their necessary and a priori status as normative facts that one must respond to if rational. Presumably, from the point of view of our fitness we could have had different moral intuitions for which we would be able to find some reasons that would justify them. Hence, even if there were some normative facts, the possibility of rationally justifying our moral judgments does not indicate the link between these facts and the truth of these judgments, which indicates that the problem of massive coincidence still holds, and that Parfit's idea of justificatory rationalism cannot avoid problems of contingent justification.

Therefore I conclude that Smith's and Parfit's conceptions of justificatory rationalism do not provide plausible views about the nature of morality in the face of empirical theories and data.

CONCLUSION

In this thesis my aim was to examine the impact of recent discoveries and developments in empirical moral psychology on thesis, conceptions and aspirations of moral rationalism. In order to assess the ideas contained among moral rationalists I followed Nichols (2004) and Joyce (2008) in dividing moral rationalism into three distinct claims: conceptual rationalism, psychological rationalism and justificatory rationalism.

Conceptual rationalism and psychological rationalism are descriptive claims. Conceptual rationalism is a claim about moral concepts, which says that it is a conceptual truth that when one judges that it is morally right to Φ then one judges that there is a reason to Φ. In other words it says that it is a conceptual truth that moral requirements are requirements of practical rationality. Psychological rationalism is a claim about the sources of moral judgments. It says that moral judgments products of our rational capacities, where these capacities refer to higher-order cognitive abilities that are contrasted with emotion and perception. Justificatory rationalism is a normative claim about the rational foundations of morality. It is a substantive claim about moral facts, suggesting that moral facts are a priori and necessary truths grasped through operations of rationality.

In the second chapter I reported on Nichols' (2002) study in which he tested subjects' intuitions about the conceptual connection between morality and rationality. Study showed that *prima facie* conceptual rationalism is false. However, Joyce (2008) argued that Nichols' study was attacking a strawman, and that one cannot empirically test conceptual rationalism. I agreed with Nichols' that Joyce's skepticism about the possibility of testing conceptual rationalist claim is misplaced, but I agreed with Joyce that study is a inconclusive evidence against conceptual rationalism.

In subsequent chapter I examined the psychological rationalism. In order to do that I presented three influential models of moral judgment and argued that evidence shows that rational capacities are not sufficient and intuitive (unconscious) and emotional processes play a necessary role in producing moral judgment.

In the last chapter I criticized two rationalist's construals of the justificatory rationalistic claim. I argued that Smith (1994) account is not a plausible explanation of possible convergence in moral opinion. Against Parfit's (forthcoming) view that, if rational in giving moral judgments, we respond to reason-giving, necessary and causally-inert facts, I argue that even if our intuitions about moral issues were different we would be in the same position as we are now, and would probably have the same attitude to these different intuitions as we have towards actual. This consideration makes the whole issue about responding to normative facts empty, since even if there were any, we would not be able to now if we are responding to them, considering the fact that evolution of moral intuitions might have been different.

References

- Blair, R., J., R. (2000). Impaired social response reversal: A case of 'acquired sociopathy'. *Brain, 123*, 1122-1141.
- Blair, R., J., R. (2003). Facial expressions, their communicatory functions and neurocognitive substrates. *Philosophical Transaction of the Royal Society, London B.*
- Cushman, F., Young, L. & Hauser, M. (2006). The role of conscious reasoning and intuitions in moral judgment: testing three principles of harm. *Psychological Science, 17*, 1082–1089.
- Damasio, A. (1994). *Descartes' error: Emotion, reason, and the human brain.* New York: Putnam.
- Darwall, S. (1983). *Impartial reason.* Ithaca, N.Y.: Cornell University Press.
- Dwyer, S. (2008). How not to argue that morality isn't innate. In W. Sinnott-Armstrong (ed.), *Moral Psychology, Vol. 1, The Evolution of Morality: Adaptations and Innateness.* MIT Press, 407–418.
- Dwyer, S. (2009). Moral Dumbfounding and the Linguistic Analogy: Methodological Implications for the Study of Moral Judgment. *Mind & Language, Vol. 24*, pp. 274–296.
- Greene, J.D., Sommerville, R.B., Nystrom, L.E., Darley, J.M., & Cohen, J.D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science, 293*, 2105–2108.
- Greene, J. & Haidt, J. (2002) How (and where) does moral judgment work? *Trends in Cognitive Sciences, 6*(12), 517-523.
- Greene, J.D., Nystrom, L.E., Engell, A.D., Darley, J.M., & Cohen, J.D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron, 44*, 389–400.
- Greene, J., D. (2003) From neural "is" to moral "ought": what are the moral implications of neuroscientific moral psychology? *Nature Reviews Neuroscience*, Vol. 4, 847-850.
- Greene, J., D. (2008a). The Secret Joke of Kant's Soul. In W. Sinnot-Armstrong (ed.), *Moral Psychology Vol 3, The Neuroscience of Morality: Emotion, Brain Disorders, and Development*, MIT Press, p. 35-79.
- Greene, J., D. (2008b). Reply to Mikhail and Timmons. In W. Sinnot-Armstrong (ed.), *Moral Psychology Vol 3, The Neuroscience of Morality: Emotion, Brain Disorders, and Development*, MIT Press, 105-117.

- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review. 108*, 814-834.
- Haidt, J. (2007). The New Synthesis in Moral Psychology, *Science* 316, 998.
- Haidt, J. & Bjorklund, F. (2008). Social Intuitionists Answer Six Questions about Moral Psychology. In W. Sinnot-Armstrong (ed.), *Moral Psychology Vol 2, The Cognitive Science of Morality: Intuition and Diversity*, MIT Press, 181-217.
- Hauser, M. (2006). *Moral minds: How nature designed our universal sense of right and wrong*. New York: Harper Collins.
- Hauser, M., D., Young, L., Cushman, F. (2008). Reviving Rawls's Linguistic Analogy: Operative Principles and the Causal Structure of Moral Actions. In W. Sinnot-Armstrong (ed.), *Moral Psychology Vol 2, The Cognitive Science of Morality: Intuition and Diversity*, MIT Press, 107-143.
- Hume, D. (1739-40/2003). *A Treatise of Human Nature*, Dover Publications, Inc., Mineola, New York.
- Joyce, R., (2008). What Neuroscience Can (and Cannot) Contribute to Metaethics. In W. Sinnot-Armstrong (ed.), *Moral Psychology Vol 3, The Neuroscience of Morality: Emotion, Brain Disorders, and Development*, MIT Press, 371-394.
- Kant, I. (1785/1964). *Groundwork of the Metaphysics of Morals*, New York: Harper Torchbooks.
- Kant, I. (1788/1993). *Critique of Practical Reason* (3rd ed.), New York: Maxwell Macmillan International.
- Kiehl, K., A. (2008). Without Morals: The Cognitive Neuroscience of Criminal Psychopaths. In W. Sinnot-Armstrong (ed.), *Moral Psychology Vol 3, The Neuroscience of Morality: Emotion, Brain Disorders, and Development*, MIT Press, 166-171.
- Knobe, J. (2006). The concept of intentional action: a case study in the uses of folk psychology. *Philosophical Studies* 130: 203-231. Reprinted in J. Knobe and S. Nichols (eds.), e*xperimental Philosophy*, Oxford University Press, 2009, p. 129-147.
- Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M. D. and Damasio, A. (2007). Damage to the prefrontal cortex increases utilitarian moral judgment. *Nature, 446*, 908–911.

- Koenigs, M. & Tranel, D. (2007). Irrational Economic Decision-Making after Ventromedial Prefrontal Damage: Evidence from the Ultimatum Game. *The Journal of Neuroscience, 27*(4), 951-956.
- Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization, in D. A. Goslin (Ed.), *Handbook of socialization theory and research* (pp. 347-480). Chicago: Rand McNally.
- Kohlberg, L. (1984). *The psychology of moral development: Moral stages and the life cycle.* San Francisco: Harper & Row.
- Korsgaard, C. (1986). Skepticism about practical reason. *Journal of Philosophy* 83:5–25.
- Korsgaard, C. (1996). *The Sources of Normativity*, Cambridge: Cambridge University Press.
- Korsgaard, C., (2008). *The Constitution of Agency*, Oxford: Oxford University Press.
- Mallon, R. (2008). Reviving Rawls's Linguistic Analogy Inside and Out. In W. Sinnot-Armstrong (ed.), *Moral Psychology Vol 2, The Cognitive Science of Morality: Intuition and Diversity*, MIT Press, 145-155.
- Mikhail, J. (2008). Moral Cognition and Computational Theory. In W. Sinnot-Armstrong (ed.), Moral Psychology Vol 3, *The Neuroscience of Morality: Emotion, Brain Disorders, and Development*, MIT Press, 81-91.
- Nado, J., Kelly, D. & Stich, S. (2009). Moral Judgment. In J. Symons & P. Calvo (eds.), *The Routledge Companion to Philosophy of Psychology*, Routledge, Taylor & Francis Group.
- Nichols, S. (2002). Is it irrational to be amoral? How psychopaths threaten moral rationalism. *The Monist* 85:285–304.
- Nichols, S. (2004). *Sentimental rules: On the natural foundations of moral judgment,* New York: Oxford University Press.
- Nichols, S. (2008). Moral Rationalism and Empirical Immunity. In W. Sinnot-Armstrong (ed.), Moral Psychology Vol 3, *The Neuroscience of Morality: Emotion, Brain Disorders, and Development*, MIT Press, 396-407.
- Nisbett, R., E. & Wilson, T., D. (1977). Telling more than we can know: Verbal reports on mental processes, *Psychological Review, 84,* 231-259.
- Parfit, D., *On What Matters*. Oxford: Oxford University Press, forthcoming.

- Piaget, J. (1965). *The moral judgment of the child* (M. Gabain, Trans.). New York: Free Press. (Original work published 1932)
- Prinz, J., J. (2006). The emotional basis of moral judgments. *Philosophical Explorations, 9*, 29–43.
- Railton, P. (2008). Naturalism Relativized? In W. Sinnott-Armstrong (ed.), *Moral Psychology, Vol. 1, The Evolution of Morality: Adaptations and Innateness*. MIT Press, 37- 44.
- Scanlon, T. (1998). *What We Owe to Each Other*, Harvard University Press.
- Schnall, S., Benton, J. & Harvey, S. (2008). With a Clean Conscience: Cleanlines Reduces the Severity of Moral Judgments. *Psychological Science, 19*, 1219-1222.
- Smith, M. (1994). *The Moral Problem*, Oxford University Press.
- Wheatley, T., & Haidt, J. (2005). Hypnotically induced disgust makes moral judgments more severe. *Psychological Science, 16*, 780–784.