

# Is Evidence from Social Psychology and Neuroscience Relevant to Philosophical Debates in Normative Ethics?

Boris Rähme

*Abstract*—This article presents some considerations concerning the relevance of empirical data and theories from neuroscience and social psychology for philosophical debates in normative ethics. While many authors hold that there are findings and theories from those fields that are relevant to normative ethics, it often remains unclear precisely how this relevance relation is to be construed and spelled out. In what follows, I critically discuss various proposals, which have recently been made in this regard by philosophers, psychologists, and neuroscientists.

Philosophical debates in normative ethics, it is often claimed, should be attentive to empirical results obtained in social psychological studies of moral cognition and in neuroscientific investigations of the biological underpinnings of moral judgement formation. In what follows I critically discuss various proposals which have recently been made in this regard by philosophers, psychologists and neuroscientists. Normative ethics, social psychology of moral cognition and neuroscientific moral psychology respond to different questions, respectively. When pursued in philosophy normative ethics is primarily concerned with the search for and assessment of arguments that lend support to, or speak against, the validity of moral statements of varying generality (principles, norms, maxims, value statements, etc.). Ethicists working in this field try to offer justified answers to questions of the following kind: why should we accept/reject the moral statement  $P^1$ ? The 'why', in this case, asks for epistemic reasons and, consequently, the expression 'should' denotes not some ethical but an epistemic obligation. Social psychologists concerned with moral cognition<sup>2</sup>, on the other hand, are primarily interested

---

<sup>1</sup> Where  $P$  might be some moral principle, norm, maxim or evaluative statement.

<sup>2</sup> Moral cognition, in the intended sense, comprises moral emotions, taken as responses to particular situations. On a wide reading of 'cognition' in terms of the processing of information,

in individuating and explaining regularities and patterns in the actual moral-cognitive practices of populations. They aim to answer questions of the following kind: which ethically normative statements do we (explicitly or implicitly) accept/reject and—more importantly—why do we accept/reject them? In this case, the ‘why’ does not ask for epistemic justifications of the propositional contents of ethical statements that we accept or for epistemic refutations of the ones we reject. Rather, what it demands is an explanation, in terms of cognitive and psychological functions, of the fact that we accept the ones we do accept and reject the ones we do reject. Such explanations may or may not be relevant to the epistemic standing of our moral beliefs and cognitive dispositions more generally. Neuroscientific moral psychology, finally, can be seen as complementing social psychological explanations by investigating the neural correlates, underpinnings and realizations of the functions involved in moral cognition. The questions that stand in its centre are of the following stripe: how are moral-cognitive functions realized by the brain and what happens on the level neural activity when we engage in moral-cognitive tasks? Again, answers to these questions may or may not bear on the epistemic credentials of human moral cognition. Subscribing to the claim that normative ethics, social psychology and neuropsychology aim to answer distinct questions does not prejudge the issue of whether data and theories from the latter two can be relevant to debates in normative ethics.

Some neuroscientists and neurophilosophers might take my description, in terms of ‘complementing’, of how neuroscientific moral psychology relates to social psychology as too tame. They would prefer to talk of reduction or even elimination instead<sup>3</sup>. A serious discussion of the various issues connected to neuroscientific reductionism and eliminativism would require far more space than this article permits. In what follows I remain neutral on these issues and follow the more conciliatory complementary view on the relation between social psychology, neuropsychology and, for that matter, normative ethics. A last preliminary point: for present purposes I take it for granted that contemporary methods in neuroscience (in particular those based on functional magnetic resonance imaging

---

moral emotions can clearly be qualified as an at least partly cognitive matter. See J.D. Greene et al., *The Neural Bases of Cognitive Conflict and Control in Moral Judgment*, in «Neuron», 44, 2004, pp. 389-400. Talk of moral cognition in this broad sense does not *per se* commit one to accepting one or another form of what is called ‘cognitivism’ in metaethics.

<sup>3</sup> See, for instance, J. Bickle, *Philosophy and Neuroscience: A Ruthlessly Reductive Account*, Dordrecht 2003.

(fMRI)) and in social psychology (in particular those based on survey techniques) produce reliable and well justified results<sup>4</sup>. My question is not whether the social psychological and neuroscientific results appealed to in the debate are reliable. Rather, I assume them to be reliable and then ask how they can be supposed to bear upon questions in normative ethics. Given the background assumptions concerning normative ethics sketched above, I take this to boil down to the question of whether and how social-psychological and neuroscientific data and theories can serve as premises (or as justifications of premises) in arguments that aim to give epistemical support to or to undermine normative ethical statements.

### 1. *Neuroscientific moral psychology and naturalized ethics*

According to a widely accepted reading of the term «neuroethics», first explicitly proposed by Adina Roskies, the field of neuroethics comprises the neuroscience of ethics and the ethics of neuroscience<sup>5</sup>. While the first branch is dedicated to investigating the neural correlates and underpinnings of ethical judgment, decision-making and cognition more generally, the second branch is concerned with the ethical problems arising within and from neuroscientific practice, i.e. with questions regarding, for instance, ethical constraints on experiment design, the ethically legitimate application of the results of neuroscientific research in various social and technological settings, and the communication of those results to non-expert audiences. However, even a cursory glance at the current literature published under the heading «neuroethics» suffices to notice that something important is left out by this *prima facie* plausible and certainly useful characterization of the field: neuroscientific ethics. Some philosophical ethicists and some neuroscientists investigating the neural correlates of moral cognition and behavior aim to draw ethically normative conclusions from neuroscientific research findings, and some even pursue the project of developing a normative ethics that is grounded in or based upon neuroscientific

<sup>4</sup> For a well-argued cautionary note on how (not) to interpret and evaluate fMRI data, see A. Roskies, *Neuroimaging and Inferential Distance*, in «Neuroethics», 1, 2008, pp. 19-30. For recent constructive criticisms of the present methodological state of survey-based experimental philosophy see R.L. Woolfolk, *Experimental Philosophy: A Methodological Critique*, in «Metaphilosophy», 44, 2013, pp. 79-87, and D. Rose - D. Danks, *In Defense of a Broad Conception of Experimental Philosophy*, in «Metaphilosophy», 44, 2013, pp. 512-532.

<sup>5</sup> A. Roskies, *Neuroethics for the New Millenium*, in «Neuron», 35, 2002, pp. 21-23.

research<sup>6</sup>. This ambitious project does not seem to be covered by Roskies' two-fold characterization of the field of neuroethics. Yet, it certainly merits being included. The idea behind the research program that I have just labelled 'neuroscientific ethics' is usefully summed up by the neuroscientist William D. Casebeer. He casts his characterization in terms of naturalized ethics but clearly intends a naturalization of the neuroscientific kind:

«The goal of naturalized ethics is to show that norms are natural, and that they arise from and are justified by purely natural processes»<sup>7</sup>.

The fact that the expressions 'arise from' and 'justified by' occur so close to each other in this statement is likely to ring an alarm bell in many readers. And, in fact, Casebeer immediately goes on to address the obvious philosophical objection that is bound to be raised against any such attempt to justify moral norms and values appealing exclusively to natural processes, i.e. the naturalistic fallacy objection:

«If this can be done», i.e. if ethics can be naturalized, «then the naturalistic fallacy is not actually a fallacy (it merely amounts to saying that you don't have a good naturalized ethical theory yet)»<sup>8</sup>.

But this does not seem to take the sting out of the naturalistic fallacy objection. Of course, if ethics can be naturalized, then the naturalistic fallacy objection is spurious, i.e. that which it refers to as fallacious is, in fact, not a fallacy at all. But given that it is precisely the impossibility of naturalizing normative ethics that the naturalistic fallacy objection purports to establish, saying the latter gets very close to saying that if normative ethics can be naturalized, then normative ethics can be naturalized. It is difficult to see how this tautology could be reasonably taken to address and counter the naturalistic fallacy objection. I am not claiming that the objection cannot be countered, that it is decisive or conclusive. My point is, rather, that the way in which Casebeer tries to counter it is ineffective<sup>9</sup>.

<sup>6</sup> See, for instance, P. Smith Churchland, *Braintrust. What Neuroscience Tells Us about Morality*, Princeton 2011.

<sup>7</sup> W.D. Casebeer, *Moral Cognition and Its Neural Constituents*, in «Nature Reviews Neuroscience», 4, 2003, pp. 841-847.

<sup>8</sup> *Ibidem*, p. 843. The no-ought-from-is objection and the naturalistic fallacy objection are distinct. In the present context, however, this distinction can be neglected.

<sup>9</sup> See, however, the more sustained attempt at countering the naturalistic fallacy objection in W.D. Casebeer, *Natural Ethical Facts. Evolution, Connectionism, and Moral Cognition*, Cambridge MA 2003, chapter 2.

Casebeer takes Millian utilitarianism, Kantian deontology and Aristotelian virtue ethics to be the «three main classic moral theories in the Western tradition», and he suggests that currently available results from neuroscientific moral psychology lend defeasible support to the claim that «the moral psychology required by virtue theory is the most neurobiologically plausible»<sup>10</sup>. But Casebeer seems to take a further step. He (implicitly, at least) takes the neurobiological vindication of the moral psychology that he assumes to be required by largely Aristotelian virtue ethics to carry over to an argument that lends defeasible epistemic support to Aristotelian virtue ethics proper<sup>11</sup>. That this is the way in which Casebeer intends his line of argument to be understood is suggested by his programmatic claim that in an appropriately naturalized ethics moral norms and values would be ‘justified by purely natural processes’. Making the argument that can be gathered from Casebeer’s remarks explicit, we get, roughly, the following:

#### Argument 1

- |      |     |   |                   |
|------|-----|---|-------------------|
| 1    | (1) | Normative ethical theories imply (or are associated with or require) <sup>12</sup> empirical predictions about the ways in which we <i>de facto</i> engage in moral cognition.  | premise           |
| 2    | (2) | If (1), then the following holds: if the moral psychological predictions implied by (etc.) a normative ethical theory T are consistent with our best neuroscientific theories of how we <i>de facto</i> engage in moral cognition, then T enjoys defeasible epistemic support from our best neuroscientific theories. | premise           |
| 1, 2 | (3) | If the moral psychological predictions implied by (etc.) a normative ethical theory T are consistent with our best neuroscientific theories of how we <i>de facto</i> engage in moral cognition, then T enjoys defeasible epistemic support from our best neuroscientific theories.                                   | from (1), (2), MP |
| 4    | (4) | The moral psychological predictions implied by (etc.) Aristotelian virtue ethics are consistent with our  | premise           |

<sup>10</sup> *Ibidem*, p. 841.

<sup>11</sup> This, at any rate, is how the moral psychologist and philosopher Joshua Greene interprets Casebeer’s line of argumen—in order to then diagnose «a *non sequitur*» along the lines of the naturalistic fallacy objection. See J.D. Greene, *From Neural ‘is’ to Moral ‘Ought’: What are the Moral Implications of Neuroscientific Moral Psychology?*, in «Nature Reviews Neuroscience», 4, 2003, pp. 847-850, 847. See also W.D. Casebeer, *Neurobiology Supports Virtue Theory on the Role of Heuristics in Moral Cognition*, in «Behavioral and Brain Sciences», 28, 2005, pp. 547-548.

<sup>12</sup> In the remainder of the argument the qualification in brackets is abbreviated by ‘(etc.)’.

best neuroscientific theories of how we *de facto* engage in moral cognition.

- 1, 2, 4 (5) Aristotelian virtue ethics enjoys defeasible epistemic support from our best neuroscientific theories. from (3), (4), MP

Having reached the intermediate conclusion on line (5), the reasoning implicit in Casebeer's remarks can be interpreted as proceeding in the following way:

- 6 (6) If some normative ethical theory  $T_1$  does, while some other normative ethical theories  $T_2, T_3, \dots, T_n$  do not, enjoy epistemic support from our best neuroscientific theories, then *ceteris paribus*  $T_1$  is in better epistemic standing than  $T_2, T_3, \dots, T_n$ . premise
- 7 (7) Neither Millian utilitarianism nor Kantian deontology enjoy epistemic support from our best neuroscientific theories. premise
- 1, 2, 4, 7 (8) Aristotelian virtue ethics does, while Millian utilitarianism and Kantian deontology do not, enjoy epistemic support from our best neuroscientific theories. from (5), (7), conjunction introduction
- 1, 2, 4, 6, 7 (9) Aristotelian virtue ethics is in better epistemic standing than Millian utilitarianism and Kantian deontology, respectively. from (6), (8), MP

Now, the argument thus reconstructed is valid. The conclusions on lines (5) and (9) do indeed follow from their respective premises. But I think that there are very strong reasons to reject the argument's soundness. Its premises are highly problematic. In particular, premise (1) is dubious. Why should normative ethical theories imply or require predictions about how we *de facto* engage in moral cognition? After all, their systematic goal is neither to describe how we *de facto* engage in moral cognition nor to explain why our *de facto* moral cognition is the way it is. Their point is, rather, to give a justified answer to the question of what criteria, principles, norms or value considerations should guide us in our moral cognitive tasks, i.e. their goal is to offer an epistemically principled answer to the question of what is the right way to engage in moral cognition. At this point the normative-descriptive and the is-ought distinctions recur. If that is correct, then premise (1) is false—normative ethical theories do not imply or require predictions about our actual ways in moral cognition—and thus both the intermediate conclusion on line (5) and the main conclusion on line (9) are shown to depend upon at least one false premise<sup>13</sup>. This diagnosis, of course, remains in

<sup>13</sup> Compare the dependency numbers on the left-hand sides of lines (5) and (9).

place even if we grant the premises on lines (2), (4), (6), and (7) for the sake of argument.

With regard to the question of how neuroscientific data and our best neuroscientific theories can bear upon debates in normative ethics, the idea behind Argument 1 seems to be this: they can offer defeasible epistemic support to and defeasibly undermine the epistemic standing of normative ethical theories. The cross-theoretical inferential link that is needed to get this idea off the ground is then introduced via the assumption that normative ethical theories imply or—in some logically less demanding sense—require predictions about the ways in which human beings actually engage in moral cognition. Since human moral cognition depends on neural realization, knowledge about how moral cognition is realized in the human brain may be consilient or inconsilient with those predictions and consequently, or so the thought seems to go, may support or undermine the epistemic standing of the normative theories from which those predictions derive. The most obvious problem with this interesting approach is that ethicists are not committed to making empirical predictions. Normative ethical theories neither imply nor, in some looser logical sense, require the truth of specific empirical predictions concerning our actual ways of engaging in moral cognitive tasks. In other words, the cross-theoretical inferential link that the reconstructed argument intends to exploit does not seem to exist. This, of course, is not to exclude the possibility of there being other cross-theoretical inferential links between normative ethics and neuroscientific moral psychology.

What, presumably, should be taken to go along with normative ethical theories is an adequacy constraint along the following modal lines: if the moral cognitive processes that would be required in actually following some given normative ethical theory T are biologically impossible for human brains to realize, then T cannot be correct. In other words: if it is biologically impossible for human brains to support moral cognition in the way that some normative ethical theory T claims to be the right way, then this speaks against T. I think that it is this point that Joshua Greene is after when he addresses the following question to Casebeer:

«So long as people are capable of taking Kant's or Mill's advice, how does it follow from neuroscientific data—indeed how could it follow from such data—that people ought to ignore Kant's and Mill's recommendations in favour of Aristotle's?»<sup>14</sup>.

<sup>14</sup> J.D. Greene, *From Neural 'Is' to Moral 'Ought'*, p. 847. It should be stressed that Greene's question is somewhat unfair, because Casebeer nowhere claims that people should ignore Kant and Mill in favour of Aristotle.

Kantian deontology and Millian utilitarianism are not so *outré* as to require, for their being applied and followed as guides in moral decision making, cognitive processes whose neural realization is a matter of biological impossibility for human brains; and, to my knowledge at least, no neuroscientist has ever claimed to have produced evidence to the contrary.

## 2. Moral judgments and post-hoc rationalizations

It merits emphasis that *mutatis mutandis* the points made above with respect to Casebeer's claims concerning the epistemic relevance of neurobiology to normative ethics carry over to a line of discussion that has been initiated by the social psychologist Jonathan Haidt. In his still immensely influential article «The Emotional Dog and Its Rational Tail» Haidt argues that, contrary to traditional «rationalist models», moral judgments are (in the majority of cases) not the result of conscious moral reasoning but of intuitive and direct responses to moral challenges<sup>15</sup>. According to Haidt's social intuitionist approach to moral judgment, moral reasoning comes into play, if at all, in the form of post-hoc rationalizations of intuitively reached moral verdicts or decisions. Haidt takes his findings to cast doubt not only on rationalist models of moral psychology but also on those approaches to normative philosophical ethics which, in his view, assign an overly prominent role to conscious moral deliberation—including Millian utilitarianism and Kantian deontology. Even though, to my knowledge, Haidt nowhere explicitly states his criticism of the «worship of reason»<sup>16</sup> in normative ethics in precisely this way, the argument on which he rests his criticism seems to be the following:

### Argument 2

- |   |     |  |         |
|---|-----|--|---------|
| 1 | (1) | Normative ethical theories imply (etc.) empirical predictions about the ways in which we <i>de facto</i> engage in moral cognition.                                  | premise |
| 2 | (2) | If (1), then the following holds: if the moral psychological predictions implied by (etc.) a normative theory T are inconsistent with social intuitionist moral psy- | premise |

<sup>15</sup> J. Haidt, *The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment*, in «Psychological Review», 108, 2001, pp. 814-834.

<sup>16</sup> *Ibidem*, p. 815.



		chology, then the latter defeasibly undermines the epistemic status of T	
1, 2	(3)	If the moral psychological predictions implied by (etc.) some normative ethical theory T are inconsistent with social intuitionist moral psychology, then the latter defeasibly undermines the epistemic status of T.	from (1), (2), MP
4	(4)	Millian utilitarianism, Kantian deontology and many other normative ethical theories imply moral psychological predictions that are inconsistent with social intuitionist moral psychology.	premise
1, 2, 4	(5)	Social intuitionist moral psychology defeasibly undermines the epistemic status of Millian utilitarianism, Kantian deontology and many other normative ethical theories.	from (3), (4), MP

Again, the argument is valid but not sound. The reason why it is not sound is not that the «central claim of the social intuitionist model»<sup>17</sup> of moral psychology is false—I actually guess that it is true, but that is irrelevant here—but, again, that premise (1) is false. If this is correct, then Millian utilitarianists and Kantian deontologists can—if they want to—consistently have both: they can accept Haidt’s claim that moral judgments are usually the result not of reasoning but of «quick and automatic moral intuitions»<sup>18</sup> and hold on to their preferred normative ethical theories without thereby committing themselves to an inconsistent conjunction of claims.

Of course, Haidt is perfectly aware of the fact that the central thesis of his social intuitionist moral psychology is a descriptive one:

«It says that moral reasoning is rarely the direct cause of moral judgment. That is a descriptive claim, about how moral judgments are actually made. It is not a normative or prescriptive claim, about how moral judgments ought to be made»<sup>19</sup>.

But whence, then, the polemic against philosophers for their worshipping reason in normative ethics?

<sup>17</sup> *Ibidem*, p. 817.

<sup>18</sup> *Ibidem*, p. 823.

<sup>19</sup> *Ibidem*, p. 815.

### 3. *Trolley problems and moral intuitions*

Considerable portions of the philosophical discussion on the epistemic status of our intuitive responses to moral problems have concentrated on what have come to be called «trolley problems»<sup>20</sup>. Consider the following two scenarios:

– Switch: A runaway trolley is heading at high speed towards a group of five persons working on the track that the trolley is running on. If nothing is done, then all five will be killed by the trolley. The group is unaware of the danger to their lives, but a bystander observes the situation and realizes that the only available way in which she could save the five lives is this: she could hit a nearby switch that would divert the trolley to a sidetrack where only one person is working. Diverting the trolley would kill that person and save the five. Is it morally legitimate to hit the switch?

– Footbridge: A runaway trolley is heading at high speed towards a group of five persons working on the track that the trolley is running on. If nothing is done, then all five will be killed. The group is unaware of the danger to their lives, but a bystander on a footbridge above the track observes the situation and realizes that the only available way in which she could save the five lives is this: she could push a large stranger that happens to be standing close to her off of the footbridge and onto the trolley's track. The large stranger would stop the trolley and be killed in the process. The five would be saved. Is it morally legitimate to push the stranger on the track<sup>21</sup>?

Social psychological experiments have revealed that test subjects regularly respond differently to the two situations presented. Considering Switch, a significant majority responds that it would be morally legitimate to hit the switch. In considering Footbridge, on the other hand, a significant majority responds that it would be morally illegitimate to push the large

<sup>20</sup> The introduction of trolley problems into the ethical debate is due to Philippa Foot. An important early elaboration is due to Judith Jarvis Thomson. See P. Foot, *The Problem of Abortion and the Doctrine of the Double Effect*, in «Oxford Review», 5, 1967, pp. 5-15; J.J. Thomson, *Killing, Letting Die, and the Trolley Problem*, in «The Monist», 59, 1976, pp. 204-217. The literature on runaway trolleys has since exploded.

<sup>21</sup> Note that the expression 'realizes' in the description of the scenario is intended to be *factive* in the scenario, i.e. hitting the switch and diverting the trolley, respectively, are in fact the only available ways to save the five lives in the hypothesized situations. This point needs to be held constant when, feeling that the scenarios are unrealistic, one tries to make them more realistic by adding further elements to their descriptions.

stranger on the track<sup>22</sup>. Since in both cases five lives would be saved at the cost of one, this result is somewhat surprising and, arguably, it stands in need of explanation.

Some philosophers involved in the debate claim that the best explanation of the fact that responses to Footbridge and Switch diverge in the way described involves the hypothesis that there is some morally significant difference between the two situations and that this difference is reflected by the diverging majority assessments of the moral legitimacy of what these scenarios respectively suggest as the only available way to save the five workmen's lives. Accordingly, they try to individuate that hypothesized difference and to explicitly characterize it in the form of a moral principle. The guiding assumption of this approach is that our intuitive responses to these moral dilemmas are veridical and epistemically reliable. They are assumed to track morally relevant truths in the sense of being responsive to morally significant differences between practical situations in an epistemically reliable way. If someone did come up with a sound argument, A, that had the content of this explanatory assumption as its conclusion, then A would at the same time offer *prima facie* epistemic support to the propositional contents of our moral intuitions quite generally. That would be a big and outstandingly significant argument, indeed.

The idea just outlined bears a distant but nonetheless non-accidental and instructive resemblance to the ways in which Descartes wanted to put clear and distinct (evident) ideas to work in his intuitionist epistemology for descriptive knowledge. One instructive way to construe Descartes' reasoning in this regard is the following<sup>23</sup>:

### Argument 3

- 1 (1) If I have the clear and distinct idea that *P*, then I am in a position to know that I have the clear and distinct idea that *P*. premise

<sup>22</sup> See J.D. Greene - J. Haidt, *How (and Where) does Moral Judgment Work?*, in «Trends in Cognitive Science», 6, 2002, pp. 517-523; J.D. Greene et al., *Pushing Moral Buttons: The Interaction between Personal Force and Intention in Moral Judgment*, in «Cognition», 111, 2009, pp. 364-371.

<sup>23</sup> For a detailed reconstruction of Descartes' intuition-based epistemology see B. Rähme, *Wahrheit, Begründbarkeit und Fallibilität. Ein Beitrag zur Diskussion epistemischer Wahrheitskonzeptionen*, Heusenstamm 2010, pp. 124-134. See also M. Kaplan, *It's Not What You Know That Counts*, in «The Journal of Philosophy», 82, 1985, pp. 350-363, in particular pp. 360-361: «Knowledge, for Descartes, is a state of cognition different—and, from the knower's point of view, discernibly different—from what we call 'justified belief'. According to Descartes, a known proposition leaves unmistakable psychological evidence of its being known».

- |         |     |   |                                    |
|---------|-----|---|------------------------------------|
| 2       | (2) | If I am in a position to know that I have the clear and distinct idea that <i>P</i> , then I am in a position to know that the fact that I have the clear and distinct idea that <i>P</i> gives me infallible epistemic support for the truth of <i>P</i> .   | premise                            |
| 3       | (3) | If I am in a position to know that the fact that I have the clear and distinct idea that <i>P</i> gives me infallible epistemic support for the truth of <i>P</i> , then the fact that I have the clear and distinct idea that <i>P</i> gives me infallible epistemic support for the truth of <i>P</i> . | premise <sup>24</sup>              |
| 1, 2, 3 | (4) | If I have the clear and distinct idea that <i>P</i> , then the fact that I have the clear and distinct idea that <i>P</i> gives me infallible epistemic support for the truth of <i>P</i> .   | From (1), (2) (3),<br>Transitivity |

Of course, epistemological infallibilism is today deemed unacceptable by most philosophers. In particular, no one committed to the idea that our moral intuitions track moral truth would want to claim that our moral intuitions give us infallible epistemic support for their propositional contents. However, if we set Descartes' infallibilist pretensions aside (i.e. if we employ a concept of defeasible justification) and relate what then remains of Argument 3 to particular ethically problematic contexts or situations, then what we get is pretty much the reasoning that, or so I would claim, stands in the background of the idea that moral intuitions are truth trackers. The reasoning parallels Descartes' argument in significant ways:

#### Argument 4

- |   |     |   |         |
|---|-----|---|---------|
| 1 | (1) | If we have the moral intuition that <i>P</i> concerning some particular morally problematic situation <i>S</i> , then we are in a position to know that we have the moral intuition that <i>P</i> concerning <i>S</i> .   | premise |
| 2 | (2) | If we are in a position to know that we have the moral intuition that <i>P</i> concerning <i>S</i> , then we are in a position to know that the fact that we have the moral intuition that <i>P</i> concerning <i>S</i> gives us defeasible epistemic support for the truth of <i>P</i> . | premise |
| 3 | (3) | If we are in a position to know that the fact that we have the moral intuition that <i>P</i> concerning <i>S</i> gives us   | premise |

<sup>24</sup> On this factivity claim concerning 'being in a position to know' see, for instance, G.H von Wright, *Logical Studies*, London 1957, p. 183: «When saying that it is possible to come to know the truth of a certain proposition, I may mean, by implication, that the proposition in question is true. In other words: only of true propositions may one—in this sense of 'may'—come to know the truth».

defeasible epistemic support for the truth of  $P$ , then the fact that we have the moral intuition that  $P$  concerning  $S$  gives us defeasible epistemic support for the truth of  $P$ .

- 1, 2, 3 (4) If we have the moral intuition that  $P$  concerning some particular morally problematic situation  $S$ , then the fact that we have the moral intuition that  $P$  concerning  $S$  gives us defeasible epistemic support for the truth of  $P$ . from (1), (2) (3), Transitivity

Both Argument 3 and Argument 4 aim to forge a link between a specific kind of ascertainable, propositionally contentful, psychological state ('having the intuition that  $P$ ,  $Q$ , etc.') on the one hand and the epistemic standing of the propositional contents of psychological states of that kind (' $P$ ,  $Q$ , etc.') on the other. In his methodologically solipsistic epistemology Descartes appealed to the privileged first-person ascertainability of one's own clear and distinct ideas. Just as Descartes' infallibilist epistemology, the introspective psychology that went along with it today seems highly problematic to many philosophers and psychologists. But there are, of course, alternatives to it. We can replace it, for instance, by a survey based social-psychology-*cum*-statistics approach to what we (i.e. what most people) think about specific moral problems like Footbridge and Switch, for instance. This approach takes statements of the form «We have the moral intuition that  $P$ » to be empirically testable by the methods of social psychology<sup>25</sup>.

In a recent article, Guy Kahane elaborates on this methodological approach and offers an argument to the conclusion that if our moral intuitions are *grosso modo* reliable, i.e. if the fact that we (the majority of us) have some moral intuition that  $P$  provides defeasible epistemic support for  $P$ , then empirical socio-psychological evidence concerning the question of which ethically normative statements we (explicitly or implicitly) accept is highly relevant to normative ethics. Kahane's argument for experimental ethics is largely in line with Argument 4. In fact, it starts from where the latter stops. Its first premise is just a rephrasing of the conclusion of Argument 4. At the same time it is more ambitious than Argument 4 in that it aims to extend the range of the epistemic significance of our moral intuitions from the level of our intuitive responses to particular cases to the level of normative ethical principles<sup>26</sup>:

<sup>25</sup> See J. Knobe - S. Nichols, *An Experimental Philosophy Manifesto*, in J. Knobe - S. Nichols (eds), *Experimental Philosophy*, New York 2008, pp. 3-14.

<sup>26</sup> The wording of the argument is quoted from pp. 428-429 of G. Kahane, *The Armchair and the Trolley: An Argument for Experimental Ethics*, in «Philosophical Studies», 162, 2013, pp. 421-445.

## Argument 5

- |       |     |   |               |
|-------|-----|---|---------------|
| 1     | (1) | Our moral intuitions about particular cases give us defeasible reason to believe in their contents.   | premise       |
| 2     | (2) | Our moral intuitions about particular cases track certain moral principles.   | premise       |
| 1,2   | (3) | Evidence about what moral principles our intuitions track gives us defeasible reason to believe in these moral principles.                      | from (1), (2) |
| 4     | (4) | Facts about what principles our intuitions track are empirical facts, and are therefore discoverable using the methods of empirical psychology. | premise       |
| 1,2,4 | (5) | Psychological evidence about the principles our intuitions track gives us defeasible reasons to endorse these moral principles.                 | from (3), (4) |

As it stands, Argument 5 is not valid: (3) cannot be derived from (1) and (2) in their present form, and (5) cannot be derived from (3) and (4) in their present shape<sup>27</sup>. However, I think that the argument could be made into a valid argument by making explicit and adding suppressed premises. (I will here not try to do so.) Therefore, I do not take this point to speak against Argument 5. In fact, I think that the argument, once made valid, should indeed be accepted as sound and convincing by many contemporary philosophers working in the field of normative ethics. I will come back to this shortly.

Kahane takes one of the most extensively discussed candidate principles that have been assumed to account for the diverging responses to Footbridge and Switch as a test and illustration case for his argument: the so-called Doctrine of Double Effect (DDE)<sup>28</sup>. Roughly, DDE states that it is morally illegitimate to intentionally harm an innocent person S, but morally legitimate to bring about harm to S as a foreseen but unintended consequence of acting on an intention to bring about good consequences that would outweigh the harm done to S in the course of that action<sup>29</sup>. According to advocates of the DDE-account it is precisely this difference between actually intending and merely foreseeing harm

<sup>27</sup> This is why I have not specified the rules of inference appealed to in the steps from (2) to (3) and from (4) to (5).

<sup>28</sup> See, for instance, F.M. Kamm, *Toward the Essence of Nonconsequentialism*, in A. Byrne - R. Stalnaker - R. Wedgwood (eds), *Fact and Value. Essays on Ethics and Metaphysics for Judith Jarvis Thomson*, Cambridge (MA) 2001, pp. 155-181.

<sup>29</sup> See G. Kahane, *The Armchair and the Trolley*, p. 422 and p. 424.

that is tracked by the majority intuitions concerning Footbridge and Switch. They maintain that the «DDE grounds, and thus explains, the moral difference between Trolley [i.e. Switch, B.R.] and Footbridge»<sup>30</sup>. Argument 5 purports to show that anyone who accepts this claim is thereby committed to accepting that our moral intuitions «can also give us epistemic reasons to believe in the DDE: they might be evidence for its truth»<sup>31</sup>.

Unfortunately, Kahane does not address the following familiar point. The idea that the DDE can account for the diverging intuitive responses to Footbridge and Switch runs into difficulties when test subjects' reactions to this variant of the trolley scenario are taken into account:

– Loop: Everything is as in Switch apart from the fact that the bystander realizes that the sidetrack rejoins the main track such that, if there were no large person working on it, diverting the trolley would not save the five lives. Is it morally legitimate to hit the switch?

Loop resembles Footbridge in that acting on the intention to put into practice what, respectively, is suggested as the only available way to save the five workmen would involve acting on an intention to harm a person. If it is our moral intuitions' implicit and reliable responsiveness to DDE that accounts for the diverging responses to Footbridge and Switch, then one should expect test subjects to respond to Loop and Footbridge in the same way, namely by saying that taking the respectively suggested action would be morally illegitimate. However, most persons confronted with Loop say that it would be morally legitimate to hit the switch<sup>32</sup>. And this suggests that our moral intuitions are not reliably responsive to the difference between intending and merely foreseeing harm that is expressed by the DDE—either because, even though such a difference exists, our intuitions don't track it, or because there is no such difference.

Be that as it may, Kahane does not explicitly endorse the DDE. Rather, he uses it as an illustrating case for the more general methodological point that he intends to make with Argument 5. In the explicit statement, Kahane presents his argument as supporting the conclusion that psychological evidence can give us defeasible epistemic justification to endorse moral principles. Commenting on the different steps of Argument

<sup>30</sup> *Ibidem*, p. 425.

<sup>31</sup> *Ibidem*.

<sup>32</sup> See, for instance, J.D. Greene et al., *An fMRI Investigation of Emotional Engagement in Moral Judgment*, in «Science», 293, 2001, pp. 2105-2108.

5, however, he is much more careful and (rightly) suggests that what his argument supports is not really (5) but only the following conditional conclusion: if (1), then (5)<sup>33</sup>. The argument should do a good job at convincing those who already hold that our moral intuitions have epistemic import of the claim that empirical psychology is highly relevant to normative ethics and, consequently, also of the claim that, for precisely this reason, the latter should go experimental. And Kahane is right in maintaining that there are many contemporary philosophers who are implicitly or explicitly committed to the claim that our moral intuitions have epistemic import<sup>34</sup>. However, pending independent justification of premise (1), Kahane's argument remains hypothetical. Dialectically, Argument 5 will be largely ineffective in a debate setting where premise (1) is not part of the common ground—in a debate, for instance, between the many philosophers who implicitly or explicitly accept premise (1), on the one hand, and the many others who do not, on the other. Someone holding that there are independent reasons for rejecting the claim that our moral intuitions give us defeasible epistemic justification for believing in their propositional contents, i.e. reasons for rejecting premise (1), will presumably not be moved at all by Argument 5.

A good case in point would be a debate between Kahane and Peter Singer. Singer suggests that the recalcitrant empirical data from social psychology on people's moral intuitions concerning Footbridge, Switch and Loop, should be taken as evidence that our moral intuitions regarding these practical dilemmas are inconsistent and therefore simply lack the kind of reliability that advocates of the truth-tracking account are after. More generally, Singer takes the divergent intuitive responses to Footbridge, Switch and Loop to be just one more example of how moral intuitions can lead astray quite generally. On his view, our moral intuitions are to be seen as often having a distorting rather than an epistemically beneficial effect on moral belief formation and decision making. This assessment is clearly based on the assumption that, contrary to what is suggested by the advocates of the hypothesis that our moral intuitions are truth-tracking, there is no morally significant difference between Footbridge and Switch to be found and consequently nothing morally relevant for our moral intuitions to track. Singer, of course, is convinced that diverting the trolley in Switch would be morally legitimate—maybe even obligatory. In this vein he writes:

<sup>33</sup> See, for instance, *ibidem*, p. 421 and pp. 439-441.

<sup>34</sup> See *ibidem*, p. 440.



«The death of one person is a lesser tragedy than the death of five people. That reasoning leads us to throw the switch in the standard trolley case [Switch, B.R.], and it should also lead us to push the stranger in the footbridge, for there are no morally relevant differences between the two situations»<sup>35</sup>.

Singer does not suggest that moral psychology is completely irrelevant to normative ethics. Rather, he assumes the results from social-psychological research on moral intuitions concerning Footbridge, Switch, Loop and other moral dilemmas to bear in an indirect way on debates in normative ethics. They

«do not themselves directly imply any normative conclusions, but they undermine some conceptions of doing ethics which themselves have normative conclusions. Those conceptions of ethics tend to be too respectful of our intuitions. Our better understanding of ethics gives us grounds for being less respectful of them».

To hold, as Singer does, that there are no morally significant differences between Switch, Footbridge and Loop is, of course, not to say that there are no differences at all. It might be, for instance, that there is a relevant psychological difference between the cases which can be appealed to in explaining test subjects' response-patterns concerning the scenarios. The most widely discussed approach along these psychological lines is the one proposed by Joshua Greene:

«We hypothesized that the thought of pushing someone to his death in an 'up close and personal' manner (as in the *footbridge* dilemma) is more emotionally salient than the thought of bringing about similar consequences in a more impersonal way (e.g., by hitting a switch, as in the *trolley* [Switch, B.R.] dilemma. We proposed that this difference in emotional response explains why people respond so differently to these two cases»<sup>36</sup>.

In other words, the explanatory hypothesis is this: what makes the difference in test subjects' divergent majority responses to Footbridge and Switch is the difference between harming in an up close and personal way and harming in an impersonal way. Utilizing fMRI, Greene and colleagues have put this explanatory hypothesis to empirical tests. The *rationale* of this neuroscientific approach to testing a functional-psychological hypothesis is that the hypothesis in question implies predictions

<sup>35</sup> P. Singer, *Ethics and Intuitions*, in «The Journal of Ethics», 9, 2005, pp. 331-352. For an instructive criticism of Singer's views see F. Tersman, *The Reliability of Moral Intuitions: A Challenge from Neuroscience*, in «Australasian Journal of Philosophy», 86, 2008, pp. 389-405.

<sup>36</sup> J.D. Greene, *The Secret Joke of Kant's Soul*, in W. Sinnott-Armstrong (ed.), *Moral Psychology, 3: The Neuroscience of Morality: Emotion, Brain Disorders, and Development*, Cambridge (MA) 2008, pp. 35-80, 43. Notice that this hypothesis also offers an explanation of why the majority of test subjects respond to both Switch and Loop that the respectively suggested only available way to save the five is morally legitimate.

about «what we should see going on» in test subjects' brains «while they are responding to dilemmas involving personal versus impersonal harm»<sup>37</sup>. Simplifying their rich findings, the central results of Greene and colleagues can be summarized in this way: moral dilemmas which suggest a course of action that would involve 'personal' harm (e.g. Footbridge), on the one hand, and moral dilemmas which suggest a course of action that would involve inflicting harm in an 'impersonal' way (e.g. Switch) typically trigger pronounced neural activity in different areas of the brain. Brain regions that neuroscientists take to be associated with emotional and affective response are significantly more active in test subjects confronted with personal moral dilemmas than in subjects confronted with impersonal ones. Consideration of impersonal moral dilemmas triggers pronounced activity in brain areas that are assumed to be associated with more abstract cognitive tasks (such as explicit reasoning) and lesser activity in those associated with emotional processing<sup>38</sup>. Plausibly, Greene takes these results to lend support to his psychological explanatory hypothesis.

Both Singer and Greene assume that these findings justify a sceptical stance toward the idea that our moral intuitions are truth trackers and, therefore, have consequences for the way debates in normative ethics should be pursued. With respect to the question of precisely how they suppose Greene's results to bear on the way debates in normative ethics should be pursued, they can be understood as bringing forward an argument along the following lines:

#### Argument 6<sup>39</sup>

- 1 (1) The fMRI data collected by Greene and colleagues lends support to the following claim: The best explanation of our (the majority of test subjects') intuitive responses to Footbridge, Switch and Loop and many premise

<sup>37</sup> *Ibidem*.

<sup>38</sup> J.D. Greene et al., *An fMRI Investigation of Emotional Engagement in Moral Judgment*; J.D. Greene - J. Haidt, *How (and Where) does Moral Judgment Work?*, in «Trends in Cognitive Science», 6, 2002, pp. 517-523; J.D. Greene et al., *The Neural Bases of Cognitive Conflict and Control in Moral Judgment*; J.D. Greene et al., *Pushing Moral Buttons: The Interaction between Personal Force and Intention in Moral Judgment*.

<sup>39</sup> Both Greene and Singer supplement this argument by an evolutionary explanation of our response patterns to personal and impersonal moral dilemmas, thus adding a further explanation of the majority responses to Footbridge and Switch in terms of what they claim to be morally irrelevant factors. See J.D. Greene, *The Secret Joke of Kant's Soul*, p. 43, and P. Singer, *Ethics and Intuitions*, pp. 333-337. For a critical discussion of this move see G. Kahane, *Evolutionary Debunking Arguments*, in «Noûs», 45, 2011, pp. 103-125.

other moral dilemmas is that our moral intuitions are responsive to the difference between inflicting harm in a personal way and inflicting harm in an impersonal way.

- |         |     |   |                   |
|---------|-----|---|-------------------|
| 2       | (2) | If (1), then the fMRI data collected by Greene and colleagues lends support to the claim that the best explanation of our intuitive responses to Foot-bridge, Switch and Loop (and many other moral dilemmas) is in terms of our moral intuitions' being responsive to a morally irrelevant difference <sup>40</sup> .          | premise           |
| 1, 2    | (3) | The fMRI data collected by Greene and colleagues lends support to the claim that the best explanation of our intuitive responses to Footbridge, Switch, Loop (etc.) is in terms of our moral intuitions' being responsive to a morally irrelevant difference.   | from (1), (2), MP |
| 4       | (4) | If (3), then the fMRI data collected by Greene and colleagues lends support to the claim that our having the moral intuitions that we have concerning Foot-bridge, Switch and Loop (etc.) does not give us epistemic reason for the truth of the propositional contents of those intuitions.                                    | premise           |
| 1, 2, 4 | (5) | The fMRI data collected by Greene and colleagues lends support to the claim that our having the moral intuitions that we have concerning Footbridge, Switch and Loop (etc.) does not give us epistemic reason for the truth of the propositional contents of those intuitions.  | from (3), (4), MP |
| 6       | (6) | If (5), then the fMRI data collected by Greene and colleagues undermines theories in normative ethics which work with the assumption that our having the moral intuitions that we have concerning Footbridge, Switch and Loop (etc.) gives us epistemic reason for the truth of the propositional contents of those intuitions. | premise           |
| 1,2,4,6 | (7) | The fMRI data collected by Greene and colleagues undermines theories in normative ethics which work with the assumption that our having the moral intuitions that we have concerning Footbridge, Switch and Loop (etc.) gives us epistemic reason for the truth of the propositional contents of those intuitions.              | from (5), (6), MP |

The argument is valid, so again the debate on the question of whether it shows what it purports to show can safely concentrate on the ques-

<sup>40</sup> In the remainder of the argument addition 'and many other moral dilemmas' is abbreviated by '(etc.)'.

tion of whether its premises are true. The premises it employs are very substantial, indeed. Each of them is controversial, to put it mildly. Since their discussion is beyond the scope of this paper, let me close with some observations on how the various arguments that I have reconstructed relate to each other.

#### 4. *Concluding Remarks*

The arguments reconstructed above are all indirect arguments in the following sense: they do not purport to show that some ethically normative statement follows from some statements of social psychology or neuroscientific moral psychology. Rather, they are epistemological arguments that purport to show that it is in one way or another epistemically incumbent on philosophers working in normative ethics to take into account the best available data and theories from social and neuroscientific moral psychology. Arguments 1, 2 and 5 are intended by their advocates to make a case for the claim that normative ethical theories should be coherent with our best neuroscientific or social-psychological data and theories. Argument 1 can be seen as an attempt to show that our best neurobiological theories can offer epistemic support to specific normative ethical theories. Argument 2 purports to show that moral psychology can epistemically undermine specific approaches to normative ethics. Argument 5 is to the effect that if our moral intuitions are reliable, then empirical moral psychology is highly relevant to normative ethics. Last but not least Argument 6, suggests a different account of the epistemic relevance of social and neurobiological moral psychology to normative ethics. It aims to show that certain findings from social psychology and neuroscientific moral psychology should be taken to cast doubt on the practice of treating moral intuitions as tracking the truth, a practice that is widespread among philosophers working in normative ethics<sup>41</sup>.

In a sense, the intermediate conclusion of Argument 6 on line (5) is directly opposed to the conclusion of Argument 5. If there is reason to believe that our having the moral intuitions that we have does not provide us with an epistemic justification of the propositional contents of those intuitions, then—presumably—*a fortiori* there is reason to deny Kahane's claim that «psychological evidence about the principles

---

<sup>41</sup> Argument 3 was a detour to Descartes' infallibilist epistemology and served only to prepare the interpretation of the thesis that our moral intuitions track the truth. In Argument 4 I propose a reconstruction of how advocates of that thesis intend their approach to work.

our intuitions track gives us defensible reasons to endorse these moral principles»<sup>42</sup>. Moreover, the main conclusion of Argument 6 on line (7) is, in a sense, directly opposed to Kahane's suggestion that normative ethics should go experimental. Greene and Singer can grant the more cautious interpretation of Argument 5 that I have mentioned above. The more cautious interpretation is, to repeat, that Argument 5 offers support to the claim that the many philosophers who, in their ethical theorizing, already work on the assumption that our moral intuitions *grosso modo* track ethically evaluative truths should—for the sake of methodological coherence, as it were—go experimental. But since Greene and Singer maintain a principled scepticism toward the thesis that our moral intuitions are truth trackers, they can remain unmoved by the unguarded interpretation of Argument 5 suggested by Kahane. Normative ethics should not, or so they say, go experimental precisely because there is epistemic reason to deny that our moral intuitions track moral truths. This suggests that the crucial question in the debate on the epistemic relevance of results from social psychology and neuroscientific moral psychology to normative ethics is whether or not our moral intuitions (*grosso modo*) track ethical truth. There are several attempts aimed at answering questions of the following kind: if our moral intuitions do (do not) track ethical truths, then what relevance does this point have for the debate on how theorizing in normative ethics should best be pursued? But a principled answer to the crucial question, one that would get one or the other of these attempts off the ground, does not seem to be in the offing.

---

<sup>42</sup> G. Kahane, *The Armchair and the Trolley*, p. 429.