Neuroscience and Metaethics: A Kantian Hypothesis

Roberto Mordacci
Facoltà di filosofia
Università Vita-Salute San Raffaele, Milano
mordacci.roberto@unisr.it

ABSTRACT

The interpretation of experimental data in neuroscientific research concerning moral decisions is controversial. One of the leading experimenters in the field, Joshua Greene, holds that the data show that deontological theories of morality are the expression of a confabulation which tries to give a rational justification for emotional responses. His arguments are criticized on the basis of a different interpretation of deontology. On the other hand, Marc Hauser, John Mikhail and others have proposed a research project in moral psychology called Universal Moral Grammar. This perspective is more promising as far as the normative dimension of moral judgment is concerned. Yet, it is suggested that rather than looking for the universal moral principles we should look for the (univerdsal) formal principles of morality, in a more thoroughly Kantian perspective.

0. *Introduction*

The current debate in neuroethics is particularly interesting for metaethics. Experimental findings seem to give a lot of work to those who look for a theory of morality which aims at being consistent with what we can know about ourselves.

The most disputed issue in this respect is the role of emotions in moral reasoning. Yet, there are many ways to take up this issue. One way is to try to make a picture of the *psychology* of morality, i.e. trying to assess the way emotions and reasoning are intertwined in the process of moral decision making. This is what the experimental findings about decision making seem best fitted for. Another way is to inquire into the *normative* role of emotions in moral reasoning, i.e. trying to understand if and in what measure emotions can be considered not only a *cause* of certain judgments and behaviours, but also (and mainly) a *justification* for those judgments. What is commonly known in metaethics as the issue of *moral normativity* is here receiving a new kind of evidence and argument, which seem to pose hard challenges to such well established accounts of morality as intuitionism, rationalism, sentimentalism, deontology and consequentialism.

Only a few authors in this debate keep the two ways apart. Many try to build up a case in which the empirical findings *do* constitute an argumentative basis for normative claims, especially in terms of a critical assessment of rival

theories: if a theory of moral normativity, i.e. a theory trying to account for the authority of moral judgments, is at odds with a reasonable account of moral psychology according to the most recent findings, then there is reason to consider that theory seriously flawed, at least in so far as it does not seem to be able to offer alternative accounts which can fit with the most relevant data.

A vigorous example of this kind of critique is advanced by Joshua Greene in his so far most philosophical published essay (Greene 2008, but see also Greene 2003 and Greene and Cohen 2004). Here, Greene first summarizes a wealth of experimental data, offering an interpretative framework which, he claims, remains at the level of a "strictly empirical" claim. Then, he proceeds to argue that "if these empirical claims are true, they may have normative implications, casting doubt on deontology as a school of normative moral thought" (p. 36). Greene is a skilful experimenter, and his scientific articles have offered what is probably the most cited set of experiments using moral dilemmas as a fundamental research tool (Greene et al. 2001; Greene and Haidt 2002). His attempt to derive normative conclusions from his own findings is therefore not only legitimate but extremely relevant for the moral theorist. His attempt is also a good example of a kind of literature which can legitimately be called *neuroethics*, since he develops a methodological claim as well as a substantial one, and this is what one should expect, if anything, from neuroethics. So, there is a general relevance in discussing Greene's hypotheses concerning deontology, since the issue here is wider, including some methodological and metaethical issues which situate this discussion on a more general background.

I will first summarize his arguments for justifying the normative conclusions he draws, and take into account the discussion that those claims have raised (Mikhail 2008, Timmons 2008). Then I will raise some objections and suggest a theoretical perspective in which the good reasons of some of his critiques can be accommodated in a less deflationary view of deontological thought. I will also briefly consider the Universal Moral Grammar project proposed by Marc Hauser (Hauser 2006), John Mikhail (Mikhail 2007) and others as a possible alternative to Greene's approach and suggest a slight modification of their perspective.

In a sense, I will use the distinction between descriptive and normative language as a critical frame against Greene's arguments, but this should not be understood as a straightforward application of the *is-ought* argument. I am not claiming that experimental data are irrelevant to normative theory. On the contrary, I consider empirical research in moral psychology as extremely relevant for moral theory, though not as a "groundwork" for normative claims, but as a kind of "critical test" offering reasons to exclude the empirically implausible implications of some normative theories.

1. *The essence of deontology*

In his 2008 essay, Greene argues that "it is possible that philosophers do not necessarily know what consequentialism and deontology really are" (Greene 2008, p. 37). Sure, it can be true. Philosophers tend to think that deontology and consequentialism are moral theories (rather: families of moral theories) or, to be more precise, models of moral reasoning. Even if these philosophical models do not reflect exactly what subjects can be described to do on a statistical basis (this is what science does), they offer the standard procedure of proper moral reasoning. In other words, according to the philosophers, deontology and consequentialism are *normative* images of moral reasoning, as much as logic is a normative image of cognitive reasoning which is not necessarily statistically prevailing among human subjects. Here, in the normative perspective, "normal" brain functioning in the sense of statistically prevalent is considered not equivalent to "correct thought". What we will end up with, along this line, would be a new edition of the old dispute on psychologism and the laws of logic, with the difference that this time the question is not whether the laws of logic but the laws of moral thought are derived from psychological (or, rather, neurological) facts and laws. Many think that Frege's and Husserl's arguments against psychologism were definitive, but – definitive or not – the question here is whether they are also applicable to the laws of moral thought or not.

Greene states his case defining deontology and consequentialism as "*psychological natural kinds*" i.e. "philosophical manifestations of two dissociable psychological patterns, two different ways of moral thinking, that have been part of the human repertoire for thousands of years" (pp. 37-38). He then concentrates on deontology. His claim is that "what we find when we explore the psychological causes of characteristically deontological judgments might suggest that what deontological moral philosophy really is, what it is *essentially*, is an attempt to produce rational justifications for emotionally driven moral judgments, and not an attempt to reach moral conclusions on the basis of moral reasoning" (p. 39).

This is a bold claim. Yet, on a very superficial level, one might observe that this is not really in line with what scientists in the modern tradition claim for their results: Galileo famously declared that he did not intend to "try the essences" ("*tentar le essenze*") and considered this as a fundamental methodological advice. Good science, in the Galilean tradition, is *not* the attempt to discover the essence of anything, but rather to investigate phenomena, trace regularities and formulate hypotheses which might be disconfirmed by subsequent research. Anyway, to be fair, this is not really a critique of Greene's ap-

proach: it seems clear that in this essay he is not playing the scientist's role; rather, he is legitimately doing the philosopher's job. Although he starts with saying that "philosophers may not really know what they're dealing with when they trade in consequentialist and deontological moral theories, and we may have to do some science to find out", I think that what *he* is really doing here is moral psychology *as a philosophical discipline*, using experimental data to justify theoretical conclusions concerning the essence of things. But after all, so far, so good: that's what many of us are in, doing moral theory (some philosophers would disparage the concept of essence, but I am not that kind of philosopher).

Seen from this perspective, Greene's essay is a theoretical critique of a particular philosophical tradition, based also – but not exclusively – on some empirical findings. What are the (theoretical) faults which Greene charges deontology with? Basically, the charge is that "Deontology […] is a kind of moral confabulation" (p. 63): the social and moral behaviour of people is the effect of "intuitive emotional responses" which deontology offers to rationalize *ex post*, so that our illusion to be rational agents can be reassured. So, in the end, "Deontology, I believe, is a natural "cognitive" expression of our deepest moral emotions", in the sense that it is the rather awkward attempt to transform an emotive response into a cognitive one[1]. On the other hand, according to Greene, "there is a natural mapping between the content of consequentialist philosophy and the functional properties of "cognitive" processes" (pp. 63-64). Thus, consequentialism is the properly cognitive response to moral dilemmas, while deontology is *essentially* an emotive response which pretends to be cognitive or introduces cognitive elements upon an already taken emotional evaluation.

From the point of view of metaethics, Greene draws the conclusion that some of the core premises of most versions of deontology are false: in particular, the idea that our deontological judgments rest on the basis of rationally discoverable moral truths. Since those judgments result from confabulations following deeply felt moral emotions, it is not necessary, and it is maybe false, to say that those emotions *just happen* to correspond to rational moral truths. It is more parsimonious to say that there are no such moral truths and that deontological moral judgments derive from our contingent emotional endowment, which is the result of evolution, culture, geography and personal history. As a consequence, there is no role for the universality (or the universal-

---

[1] Greene is explicit in characterizing in defining "cognitive" representations as "inherently neutral", in the sense that they do not trigger particular behavioural responses, while "emotional representations generate automatic effects" (Greene 2008, p. 40). The two kinds of representation seem to correspond to different areas in the brain (dorsolateral surfaces of the prefrontal cortex and parietal lobes for cognitive processes; amygdala and the medial surfaces of the frontal and parietal lobes for emotional processes).

izability) of deontological principles, such as, say, respect for persons. Deontological judgments can have a *general* (not a universal) value so far as some basic emotional reactions tend to be similar across cultures. But there are great differences between these reactions, especially if you look for more detailed emotional responses issuing in determinate moral judgments as, say, the legitimacy of lying or the normativity of natural heterosexual procreation. Thus, it is only contingent that we find ourselves agreeing about a certain deontological rule (e.g. do not kill innocent human beings), and this agreement should not be understood as an expression of rationality: it is simply the fact that we *just happen* share some gut reactions towards some morally relevant situations. Sure, there is the work of post-hoc rationalization, which introduces a lot of potentially critical stuff into the process of decision making. Yet, Greene's and Haidt's findings suggest (or at least can be interpreted as suggesting – see also Haidt 2001) that these rationalizing processes are substantially irrelevant for the most deontologically pure judgments, and in any case they are not the *essence* of deontological theory. Rather, deontology is essentially an emotionally driven process which its proponents disguise as a rational justification for action. This is why deontologists do not know what deontology really is. On the contrary, the only legitimate rational process concerning decision is the "cognitive", slower, calculating one which is adequately interpreted by the consequentialist tradition.

There is nothing particularly new in such an interpretation of the basis of morality. David Hume was even more radical in denying to reason *any* independent role whatsoever in moral argument. Yet, his perspective was also not one leaving *any* universalist claim aside, since he himself repeatedly declared that moral sentiments are natural, "since there never was any nation of the world, nor any single person in any nation, who was utterly depriv'd of them, and who never, in any instance, shew'd the least approbation or dislike of manners. These sentiments – Hume continues – are so rooted in our constitution and temper, that without entirely confounding the human mind by disease or madness, 'tis impossible to extirpate and destroy them" (Hume 1739-40, 474). Hume's claim is quite close to the idea of "human nature" as a kind of essence, in this case. This claim is not likely to be shared by contemporary critics of the notion of human nature and to many who consider the biological account of human life as no more than a contingent synthesis of various elements that evolution and technology might radically change in the future. This is a crucial point for moral theory, since if there is anything like a "human nature" one can raise an argument in favour of universalism even in a thoroughly sentimentalist perspective (we share some universally natural common feelings); I suspect that Greene and others in this discussion would not accept this universalist claim and would only concede that the empirical data suggest that some moral emotions seem to have a general prominence

among the investigated subjects. So far, but in the future we may have different feelings.

Also, there is nothing lethal in this attack for deontology as such: as Mark Timmons (2008) has pointed out, there might well be a sentimentalist deontology which departs from the long-standing tradition of entrenching deontological thought in pure rationality. Something similar might even be suspected about Jeremy Bentham's *Deontology* (the unfinished work to which Bentham worked sparsely from 1814 to 1831). And yet, I believe that a sentimentalist deontology would soon be asked to give up, from the normative point of view, to the rationally superior perspective of consequentialism (and in fact, what Bentham calls deontology is really the part concerning private morality of a thoroughly consequentialist system).

So, Greene's case is mainly that the empirical findings collected by him, his collaborators and a host of other researchers in the field of neurosciences constitute a powerful argument against the (prevailing) self-image of deontological theory. And this is understood as an argument in favour of consequentialism as a more powerfully cognitive attitude towards moral issues. Since this kind of attack is likely to be replicated, in various versions, it seems interesting to suggest some critical points.

First, if the empirical results do suggest that some non-consequentialist answers, such as those reported in the so-called "personal" dilemmas[2], are irreducible (that is, persons go on to refuse to directly kill one person in order to save five), this should be taken at face value and be considered as an argument *against* consequentialism. As Bernard Williams pointed out, "we are partially at least not utilitarians, and cannot regard our moral feelings merely as objects of utilitarian value. Because our moral relations are partly given by such feelings, and by a sense of what we can or cannot "live with", to come to regard those feelings from a purely utilitarian point of view, that is to say, as happenings outside one's moral self, is to lose a sense of one's moral identity; to lose, in the most literal way, one's integrity. At this point utilitarianism alienates one from one's moral feelings" (Williams 1973, pp. 103-104). So, it should sound strange that we cannot literally make rational sense of these feelings against the cost-benefit analysis: if our brain should be interpreted as having no way to give any reasonable meaning to these anti-consequentialist feelings but to say that they just happen to be produced by evolution, we should feel rather alienated at that thought. And well, this is not the way it feels: when we make deontological judgments, at least in the most obvious cases, we think that we are defending our sense of integrity rather than being possessed by some kind of external force. Of course, Greene is not arguing straightforwardly for utilitarianism against deontology, but against the self-

---

[2] Greene has become skeptical about the viability of the distinction between personal and impersonal dilemmas. See his reply to Mikhail and Timmons, 2008.

image of deontology itself; yet, he elsewhere argues (Greene 2002, Greene and Cohen 2004) that consequentialism is "the best available standard for public decision making" (Greene 2008, p. 77), so it seems that his arguments should build up at least a presumption in favour of some form of consequentialism. This presumption is indeed at work but does not really seem to be vindicated by the findings reported. On the contrary, if consequentialism is the result of detached, cold evaluation of consequences made on the basis of the "cognitive" ability of calculating some of the effects ahead of some time, we should wonder whether our moral life is best served by following this detached way of thought or by listening and interpreting some of our spontaneous inclinations.

Second, the fact that "Our most basic moral dispositions are evolutionary adaptations that arose in response to the demands and opportunities created by social life" (Greene 2008, p. 60) does not mean that they cannot count as *reasons*. To assume that any rationalization is a kind of *ex-post* confabulation seems exactly to assume that our feelings should *never* count as reasons. And in this respect, the processes under what is called S1 (fast, immediate processes) in the Double Process Theory ask for rational interpretation no less than those under S2 (slow, mediated ones). So *why call the search for a justification a confabulation*? Greene suggests that the deontologist's claim is to discover that those feelings *just happen* to correspond to eternal moral truths. But this is not an unavoidable path for deontologists: as it is easy to suggest, constructivist accounts of moral reasons can accept that we use our moral feelings as a starting point to build up a consensus over what we would propose as principles that no one could reasonably reject (cfr. Scanlon 1998). It is no use to say that this kind of process leaves the feelings just as they are (an argument which Greene calls the GIGO problem: "Garbage in, garbage out", Greene 2008b, p. 116), because the point is not to take out a principle from a feeling, but rather to validate or refute a feeling as a reason which can be justified by a principle. After the process of reflective thought, the content does not change, but its formal structure is different: what is just a feeling which I happen to have, if it can be assumed as a reason and translated into a principle which it would be unreasonable to refuse, becomes a proposed universal law, which is offered to us as rational agents to share and to live with. This is not strange at all, and we can use that principle to help our reflection in new and hard situations. Consequentialism as a psychological natural kind cannot take into account this kind of reflective work in any reasonable way, and this suggests that it cannot be considered the only "cognitive" process involved in moral reasoning.

Third and most important: psychological kinds are not normative theories, as Greene himself admits. But then, why deriving normative conclusions from the description of psychological kinds? Basically, we do moral theory because we want to know whether our common sense judgments may have some

kind of justification. But justifying is not explaining. And moral theory is not genealogy. We have a lot of morally relevant emotions, some of which do offer good reasons to act in a certain way under certain circumstances. But we do not endorse them when they do not seem to be a sufficient justification for our behaviour: what Christine Korsgaard (1996) calls "reflective endorsement" is not an attempt to substitute rational principles for natural feelings, it is rather to validate those feelings as good reasons for action. And validation is not confabulation (otherwise, science as a validation procedure would be confabulation as well). On the other hand, if the only available form of justification is the consequentialist calculus, then the real question is: if consequentialism is also just a psychological natural kind, then why should it be the only one to have normative authority – and why should we deny that the deontological psychological kind has any of that authority? Offering a genealogy of moral judgments is certainly useful for normative theory: for example, we can certainly say that a moral theory which would exclude any feeling from the exercise of moral judgment would be mistaken. The point is that no serious philosopher has ever maintained such an absurd view. Everybody knows that Aristotle defined choice as "desiring thought, or thinking desire" (*NE,* book VI, 1139 b), and Kant famously said that there is no man without moral sentiments (indeed, one who lacks or loses such a sensibility would be "morally dead") and that all we have to do is to strengthen and cultivate them (Kant 1797, 400).

When the question is "on what basis can a moral judgment be considered valid?", the point is not only whether it takes feelings into account: the point is how it manages to make the feelings good reasons for action and to separate those feelings which can be good reasons from those which cannot. The reasons take feelings into account but do not derive their authority from the feeling themselves. A good reason is a reason which can be shared *as a reason* for action. And if we try to define a principle which does not take into account those feelings, or which would suggest us that we should systematically violate our deep moral emotions, then we would have a highly implausible moral theory. Just like moral consequentialism is. As a normative theory, in fact, consequentialism is a peculiarly abstract one. As a psychological kind, it is a particularly alienating one.

Thus, maybe there are other ways to read the relevant empirical findings of neuroethics: trying to derive normative consequences from the description of abstract psychological kinds seems a debatable method.

2. *The linguistic analogy*

The battlefield of neuroethics is far from being dominated by the anti-deontological argument. From the same body of evidence, as well as from other experiments using moral dilemmas as a fundamental test for subjects taking hard decisions, a number of authors have built up a case for a straightforwardly deontological interpretation of the data (Hauser et al. 2007). Marc Hauser (Hauser 2006), John Mikhail (Mikhail 2007) and others are developing a research program of great complexity, which aims at giving a general framework for the interpretation of moral thought along the lines of an analogy with language. John Rawls had already devised such an analogy in a passage from *A Theory of Justice* (1971) and the authors are trying to bring it forward and to give it a basis in empirical evidence. The fact that in linguistics the Chomskyan project of entrenching a universal grammar on a scientific basis has gained prominence (cfr. Moro 2008) offers a good example for connecting empirical data with a general speculation on *universal,* and not just general, features of the cognitive faculties. This research program is called Universal Moral Grammar (UMG).

The collision of the UMG model with the sentimentalist account of deontological rules is direct and challenging: the empirical data show some differences in the responses to various kinds of dilemmas, which might suggest that there is an underlying structure of rules which is replicated across individuals and groups. This recurrence does not depend on the activation of the same cortical areas, but on the influence of recurring patterns of reasoning; so it is suggested that the relevant factor in deliberation is a cognitive, though not conscious, structure and that in this respect computational theory is at least as relevant as the use of fMRI.

The hypothesis behind the UMG project can therefore be interpreted as a defence of a cognitive account of moral thought, but in a different sense of the term "cognitive". While Greene uses it as a way of indicating the processes which are attributed to the system of slow and mediated mental functions (S2 in the Dual Process Theory), the account in the UMG project is rather that of an underlying structure which is the source of some aspects of both kind of systems (S1 and S2) and so accounts for both the emotive and the "cognitive" aspects of moral thought. If the analogy with linguistics holds, the idea is that just like the basis of language are embedded in our biological and mental endowment as human beings (as we have evolved so far), generating the variety of the natural languages on the basis of an invariable structure, so the fundamental principles of morality constitute a kind of natural framework upon which the historical moralities have been generated in a continuous interplay with the natural and cultural environment. To state it boldly: if there is a logic behind the variety of natural languages, there can be a logic behind the variety of historical moralities. Therefore, the research program looks for these structures, in the form of some recurring principles, such as the duty

against intentional battery and, quite surprisingly, the principle of double effect.

Now, to put it simply, the idea of a Universal Moral Grammar might be considered a new and extremely refined version of an *intuitionist* model of moral thought, where intuitions are cognitive acts or structures and not just emotions. In this perspective, intuitions are the basis of our moral reasoning: some fundamental principles are at least implicitly present in our mind, offering the foundations for a number of less abstract and more determined rules which may differ between the different cultures along parameters. In the UMG project, the interesting advancement in comparison with the old-fashioned intuitionism of, say, W. David Ross is in the idea, drawn from both linguistics and the Rawlsian constructivist model, that the fundamental principles are to be understood more as structures of moral thought than as self-evident "*prima facie* duties". The findings suggest that at least one principle (the principle of double effect) "may be operative in our moral judgments but not open to conscious introspection" (Hauser et al. 2007, p. 1): its influence is found in subjects of different cultures and with different educational level, although the subjects themselves do not seem to be aware that their judgments in the case, say, of the footbridge dilemma, are driven by that principle. In a more general perspective, Hauser suggests that "moral judgments are mediated by an unconscious process, a hidden moral grammar that evaluates the causes and consequences of our own and others' actions" (Hauser 2006, p. 2).

This perspective is ambitious and has a very wide scope. It does not need to go into the vexed question of the emotional or rational foundation of morality, since its basis is not thought to be an expression only or mainly of one of the two psychological processes: the Universal Moral Grammar is an underlying set of operating principles which are translated into reflection together with emotions. The advantage point is that if the basis of morality is thought to be a *process* rather than a faculty (emotion or reason), and if I can find the operating laws of the process, I might be able both to understand the subsequent structures underlying the cognitive processes and the deviation from the standard procedure.

Yet, I think that this project has not developed this intuition to its full strength. The ambiguity of this description is the fact that the underlying principle(s), though the subjects seem not to be aware of it (them), can be expressed in a highly refined manner in a moral theory, just like the one that used to support the principle of double effect (PDE) in the classical Thomist perspective. PDE does not seem to be totally unconscious and unaccessible to awareness, if some ingenious philosophers can devise a definition of it and pose the conditions under which the principle is thought to work. So one wonders: is the underlying principle really "inaccessible to conscience" or is it just *im-*

plicit and *open* to reflective scrutiny, so that an adequate moral theory can bring it to the fore?

Another ambiguous point is the status of the underlying principles we should look for: are they *material* principles, indicating determinate behaviour patterns (do not kill, do not lie, show gratitude, tell the truth and so on) or are they one or more *formal* principles (just like the principle of double effect is), i.e. rules indicating not *what* we should want to do, but *how* we should want to do anything we want to do? As I will now suggest, this is exactly what makes the difference between an intuitionist approach and a *transcendental* one, the one Kant suggested on the basis of the absolute formality of the fundamental principle of morality (which is not a determinate duty, but the rule of all valid duties).

3. *A Kantian hypothesis*

One of the dimensions of the issue at stake here is the need to have an adequate theory of normativity: if we are not clear on what is a reasonable account of the authority which moral judgments claim to have, we may end up conflating different levels of discourse and asking the wrong questions to experiments and theories as well.

My suggestion to experimenters, in a very humble tone, is that they might consider the possibility that, after all, Kant was right in believing that the foundations of morals are exclusively *formal*. The Categorical Imperative is *not* a moral principle, it is the principle of morality: it does not prescribe what to do but how a reason which pretends to justify an action has to be structured. In this perspective, it is not so much a matter of finding the commonly shared moral principles by way of a survey, but of finding the recurring formal structures underlying the different moral principles and rules that the various cultures have devised. Therefore, if we make an extensive survey concerning the moral opinions of people (like the "Harvard moral sense test"), we will probably not find that principles as "do not lie" or "hold promises" are universal *as such*. Yet, we might find that the process through which the subjects elaborate reasons in order to justify their actions can include those principles but in the end depend on some *formal* constraints, of which those principles are just some of the possible material versions. The rules concerning lying, promising and benefitting others, for example, may depend on the culture and on the history of the acting subject. But what makes the subject able to *reflect* upon what he is doing (while he pretends to be able to justify it) is the operation of a kind of law which does not contain a determinate moral content ("do not kill the innocent") but which imposes an internal requirement to any working

(i.e.: justifying) practical reason ("you cannot will as a universal law the reason you are pretending to act on").

In this perspective, duties might be thought as the result of a complex process in which inclinations, desires and external conditions create the basis of an action which the agent wants to perform. Within this act of willing, the agent has a number of reasons which interact with each other. Some of these reasons pretend to be justified and justifying, i.e. they claim to have a normative role. What makes one or more of those reasons really justifying is not their conformity to the *content* of a moral principle, but the possibility of willing that content (whatever it is) as a universal law. If willing that content as a universal law brings me to a practical contradiction (a contradiction in the will), then that reason is not a (morally) good reason. This is, basically, the Kantian suggestion concerning the foundations of moral normativity. If this is right, we should be able to find some effects of this underlying structure, although it needs not being *de facto* known to everyone who acts morally. Since it is a formal structure of practical thought, it should be found virtually everywhere and operating also underneath the level of consciousness, but I see no reason why it should be impossible to make it sufficiently clear upon reflection. There are many ways of expressing that kind of formal operating rule, which may retain a very general resemblance even under different definitions ("respect for persons", the Golden Rule, the rule against intentional battery, even the principle of double effect, and, not surprisingly, even a possible interpretation of the utility principle). What makes these formulations similar is their formal status of laws of a rational will, which in the end might be reduced to an extremely general formal principle: what may be called the principle of *non contradiction in the will*. So, rather than material principles we should be able to find recurring logical structures of practical reasoning, which do not constitute moral principles but rules of practical thought. Moral principles inscribe the moral sentiments into the logical structure of practical thought. And this might account for the recurrence of some contents of morality which seem to be shared, to a certain extent, through cultures: it is extremely hard to will as a universal rule certain kinds of action, at least in normal circumstances. The rest is up to the virtually infinite variety of interpretations of the relation between feelings and practical thought, along lines which in the history of moral philosophy were made famous by many great thinkers such as Aristotle and Kant.

In this perspective, consequentialism seems to be only a part of the story: of course we take into account the consequences and we do compare more versus less good or evil. But that's not the only way our practical thought judges. Among its rules, the universal principle of not willing what cannot be willed without contradiction offers a foundations for many different duties and reasons for action. The consequentialist calculus comes in when I recognize that

an action (e.g. the trolley dilemma in its simple form) is not an example of using someone merely as a means, and that therefore, in those situations, what happens is more the result of a tragic situation than the object of a choice directed at killing someone. The essential point is to keep it clear that showing how we arrive at a decision is not, not yet, showing how we are able to justify that decision. To look for the principles of justification is not always the same as looking for the processes of explanation.

References

Greene J. et al, 2001. "An fMRI investigation of emotional engagement in moral judgment", *Science* 14 september: 2105-2108

Greene, J. and Haidt, J., 2002. "How (and where) does moral judgment work?" *Trends in Cognitive Sciences*, 6(12), 517-523

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. et al, 2004. "The neural bases of cognitive conflict and control in moral judgment", *Neuron* 44, 389-400

Greene J. and Cohen J., 2004. "For the law, neuroscience changes nothing, and everything", *Phil. Trans. R. Soc. Lond.* B 359, 1775–1785

Greene J., 2008. "The secret joke of Kant's soul", in W. Sinnott-Armstrong (ed.), *Moral Psychology. Vol. III: The Neurosciences of Morality: Emotion, Brain Disorders, and Development*, MIT Press, Cambridge (Mass.), 35-79

Greene J., 2008b. "Reply to Mikhail and Timmons", in W. Sinnott-Armstrong (ed.), *Moral Psychology*, 105-117

Haidt, J., 2001. "The emotional dog and its rational tail: A social intuitionist approach to moral judgment", *Psychological Review 108*, 814-834

Hauser M., 2006. *Moral Minds: How Nature Designed our Universal Sense of Right and Wrong*, HarperCollins, New York

Hauser M.O. et al., 2007. "A dissociation between moral judgments and justifications", *Mind & Language* 22 (1), 1-21

Hume D., 1739-40. *Treatise of Human Nature*, ed. by L.A. Selby-Bigge. Vols I-II of the *Philosophical Works of David Hume*, Oxford University Press, Oxford 1978

Kant I., 1797. *Metaphysics of Morals,* ed. by M.J. Gregor, Cambridge University Press, Cambridge (Mass.) 1996

Korsgaard C.M., 1996. *The Sources of Normativity*, Cambridge University Press, Cambridge (Mass.)

Mikhail J., 2000. *Rawls' linguistic analogy: A study of the "generative grammar" model of moral theory described by John Rawls in "A Theory of Justice",* Doctoral dissertation, Cornell University, Ithaca, NY

Mikhail J., 2008. "Moral cognition and computational theory", in W. Sinnott-Armstrong (ed.), *Moral Psychology*, 81-91

Moro A., 2008. *The Boundaries of Babel. The Brain and the Enigma of Impossible Languages*, MIT Press, Cambridge (Mass.)

Rawls J., 1971. *A Theory of Justice*, Harvard University Press, Cambridge (Mass.)

Timmons M., 2008. "Toward a sentimentalist deontology", in W. Sinnott-Armstrong (ed.), *Moral Psychology*, 93-104