

---

MASSIMO REICHLIN

Università Vita-Salute San Raffaele, Milano

reichlin.massimo@univr.it

---

# THE NEOSENTIMENTALIST ARGUMENT AGAINST MORAL RATIONALISM: SOME CRITICAL OBSERVATIONS

*abstract*

---

*On the basis of the empirical evidence concerning the role of emotions in moral judgments, new sentimentalist approaches to metaethics have been proposed. Nichols' theory of sentimental rules, in particular, associates the emphasis on emotive reactions to the relevance of a normative body of rules that guide our judgment on actions. According to Nichols, the emotive mechanism of concern explains the acquisition of the moral capacity and, together with the evidence on psychopaths and autistic children, shows the implausibility of a) moral rationalism, both as a conceptual and as an empirical thesis; b) motivational internalism; and c) moral objectivism. However, if we distinguish between i) the initial acquisition of morality in children and ii) the adult experience of it, we can see that to accept a central role of the emotive mechanisms in the first is not to have shown their centrality in the second. In particular, it is not possible to account for the normative theory in purely emotive terms, even though we accept that their emotive connotation favours the evolutive success of the norms. A moderate rationalist view, grounded on the notion of reflective endorsement and on the cooperation between emotions and rational capacities seems quite compatible with the empirical evidence and can justify plausible forms of internalism and objectivism.*

*keywords*

---

*Moral sentimentalism, moral rationalism, moral internalism, moral objectivism, Nichols.*

1. Recent research in moral psychology and on the neural bases of moral judgment has highlighted the relevance of automatic reactions of an emotive kind in human morality. This research has empirical character, *i.e.* it aims to provide a reliable description of how our moral capacity *actually* works; according to some, it also has normative relevance, that is, it suggests how we should (or should not) use our moral capacities as well (Greene 2003 and 2008; Singer 2005). To say that neuroscientific research has such a normative bearing is however highly controversial, as it seems to imply a problematic derivation of “ought” from “is”: and some commentators have rightly insisted that in order to reach any such normative conclusions we must always presuppose certain moral intuitions, while the neuroscientific results are in fact doing no real work in the argument (Berker 2009).

It is much more plausible, however, to say that this research has some direct or indirect implication of a metaethical kind, that is, on our ideas concerning the nature of ethics. Metaethical propositions are not prescriptive, but descriptive: they aim to explain what *moral* judgments are, and whether they can be said to have any kind of truth-value. According to several authors, research in moral psychology and neuroethics justifies a radical revision of our traditional image of the moral capacity: specifically, a criticism of rationalist views, according to which our moral judgments are essentially the work of reason, and a vindication of a sentimentalist approach, that considers them the expression of our sentiments and affective reactions. Some explicitly say that there is empirical evidence for an emotivist conception of morality (Haidt 2001), according to which moral judgment belongs to moral sense, that is, to “an innate preparedness to feel flashes of approval or disapproval toward certain patterns of events involving other human beings” (Haidt and Joseph 2004, 56).

Although the data clearly do not show the truth of emotivism concerning the nature of *moral language*—that is, of the view holding that moral judgments have the *logical* function to express certain peculiar emotions (Joyce 2008)—they in fact seem to count as reasons in favor of a sentimentalist approach that, both from an explanatory and from an explicitly causal point of view, attributes a central role to emotive reactions and an altogether subsidiary role to rational reflection. In particular,

Haidt's social intuitionist conception aims to show that moral judgments are caused, in the great majority of cases, by the bursting in of certain emotions, while the arguments offered in their favor are almost always *ex post* rationalizations, that is, clumsy attempts to give reasons to support our automatic responses: these attempts are generated in the context of a social request of justification and have no causal or explanatory role in generating the judgments in the first place.

This model seems to leave no space for altogether ordinary psychological processes, such as first-person solitary moral reflection and deliberation, and must consider all moral arguing as a lawyer's attempt to defend a moral truth that is simply accepted a priori. Moreover, it seems unable to account for the empirical fact that controlled "cognitive" processes are effective in (sometimes) contrasting the emotive responses, overwriting rational responses based on consequentialist computations (Greene *et al.* 2001, 2004 e 2008). In other words, by making reason a complete "slave of the passions", this emotivist view is in contrast with the empirical evidence that suggested the "dual-process model".

2. Shaun Nichols' model of the sentimental rules offers a much more complex formulation of the neosentimentalist approach: it not only establishes the disposition to feel certain emotive reactions, particularly with reference to others' sufferings, as the fundamental component of the moral capacity; it also stresses the importance of some normative perspective, that is, of a body of rules concerning the approval and disapproval of certain kinds of behavior (Nichols 2002a, 2004, 2008).

Building on the data from several psychological experiments, Nichols argues for the following points: a) morality—and particularly the capacity to make *core moral judgments* (*i.e.*, judgments that actions causing suffering to other people are wrong) and to distinguish between moral and conventional violations—is explained by the affective mechanism of *concern*; this accounts for the fact that three-year old children and autistic children are able to make core moral judgments while still lacking a developed theory of mind; b) the importance of this emotive mechanism, coupled with the empirical data concerning psychopaths and people suffering from focal lesions of the ventromedial section of the prefrontal cortex, shows the implausibility of moral rationalism; c) the fact that psychopaths often display an intact capacity for moral judgment shows that knowledge of the sentimental rules can be dissociated from the emotive mechanism: moreover, subjects suffering

from focal lesions in the prefrontal cortex seem to be the living evidence of the theoretical possibility of amoralism (Roskies 2003) and therefore show the fallacy of the conceptual internalism of the rationalistic view; d) the cultural fitness of emotively laden rules explains why certain rules do survive throughout human history and are largely accepted: this makes the objectivist explanation of moral “progress” offered by the rationalist view redundant. In other words, the partial and progressive overlap between moral codes can be explained by the similarity between our affective reactions; we simply happen to evolve such sentiments and there are no deeper moral facts that might vindicate moral objectivism.

In what follows, I will suggest some reasons to believe that the very fact that Nichols’ sentimentalist view is much more plausible and attractive than Haidt’s weakens his attack against moral rationalism, limiting its efficacy to radical or extreme forms of ethical rationalism.

3. In order to assess Nichols’ basic points, let us distinguish between i) the initial experience of morality which is found in children and which emerged phylogenetically at a certain point in the evolution of the species *Homo* and ii) the much more complex experience of morality that is found in healthy adult humans. Nichols’ research concentrates on level i): it aims to account for those basic judgments which he calls *core moral judgments*. A direct and easy way to defend moral rationalism would obviously be to deny that such judgments — that, according to many moral psychologists are found in three-year old children—are in fact authentic moral judgments. We might say that the moral faculty presupposes a much larger mastery of moral concepts and of complex moral reasoning and that, therefore, the data from moral psychology do not undermine moral rationalism. However, Nichols appropriately notes that children’s moral judgments resemble the adults’ ones in many ways; moreover, a considerable part of our daily normative experience is in fact based on those same mechanisms that are at work in core moral judgments.

It seems sensible, then, to accept these judgments as integral to the moral capacity and to find in them the essential elements contributing to moral judgments in adults. Nonetheless, the fact that core moral judgments depend on emotive reactions at level i) does not necessarily imply a sentimentalist explanation of level ii); that is, it does not demonstrate that emotive reactions to the others’ suffering are the decisive element of the moral capacity, as it is possessed by adults, or that all moral judgments

depend on the concern mechanism. In effect, although Nichols considers highly implausible that the concern mechanism should have a merely evolutionary role, so that adult moral judgment could function in its total absence, he deems just as much dubious the “on-line hypothesis”, according to which the making of a moral judgment would always presuppose the activation of the affective mechanism: thanks to the normative theory, it seems very likely that adult moral judgments can often be made without any such activation.

Thus, even accepting that the emotive mechanism of concern is essential for the starting of the moral faculty (Nichols’ point a), we need to have a clearer view of the relationship between this mechanism and the normative theory in order to see whether the model of the sentimental rules can count as a refutation of moral rationalism (Nichols’ point b). Now, both elements seem necessary, according to Nichols: the ability to feel certain emotive reactions is a necessary presupposition of the moral judgment—since individuals with congenital frontal lesions never acquire the moral capacity; but the emotive reactions are not in themselves sufficient to characterize morality and to distinguish it from other normative spheres. In fact, on the one hand, in cases of natural disasters, or of human interventions causing suffering with a view to larger benefits (for example, in the medical field), we can observe emotive reactions similar to those linked to core moral judgments, but no moral judgment is generated. On the other hand, Nichols’ experiments themselves show that prohibitions relative to disgusting behaviour, which are reinforced by strong emotive reactions, acquire the same weight of moral prohibitions, even though common sense clearly distinguishes between the two kinds of prohibitions. (Nichols 2002a). This suggests that the emotions are insufficient to distinguish the moral from the non moral sphere, and that the normative theory does not simply embody those emotive reactions. As a matter of fact, some non moral prohibitions relative to disgusting behaviour may be supported by negative emotions that are even stronger than those relative to some moral prohibitions, for example those tied to rules of justice and fairness in distribution, or those relative to promises: however, our normative theory considers the second as a clearly distinct kind of violation, and one much more serious than the first.

It seems therefore highly plausible to say that the normative theory that structures our emotive reactions reflects a relevant work of rational reflection as well. It is not the case that our set of rules simply expresses our emotions; it also reflects the work of rational reflection, a work which can partially shape

our emotive reactions. This work manipulates specifically moral concepts—not only harm and physical suffering, but also injustice and unfairness, disrespect of others' dignity, humiliation of others, and the like—and on the basis of these concepts attributes a particular degree of importance to *specifically moral* violations; the acknowledgment of the peculiar importance of moral violations, on its turn, generates new emotions and sentiments that were not involved in core moral judgments. The very fact that Nichols' theory attaches a relevant role to the normative theory, therefore, shows that sentimental reactions are insufficient in order to give a complete explanation of morality and of its distinction from other forms of normativity. If this is so, it is very likely that also in the initial acquisition of the moral capacity an essential role is played by the normative set of rules that enable to judge the appropriateness of feelings and actions; without the aid of the normative theory, children would never be able to distinguish between the moral from the non moral sphere, and would never learn the distinction between rules forbidding actions that are malicious or unfair and those forbidding actions that are simply disgusting.

4. A possible sentimentalist rejoinder is to insist that the normative theory is itself the product of our emotive reactions: it is the emotive mechanism that accounts for the rules we have, and the normative theory is simply the systematization our sentimental responses. Nichols himself proposes the “affective resonance hypothesis”, according to which the annexing of a rule into our normative system is considerably favoured by its cultural fitness, which in turn depends on our emotive reactions: that is, the more a rule is tied to a behaviour that in itself generates a remarkable emotive reaction, the more it is likely to be permanently adopted and handed down to future generations (2002b; 2004). However, we must note that the fact that being emotively reinforced favours the evolutionary success of a rule is quite compatible with the hypothesis that the normative theory is not a mere systematization of our spontaneous affective reactions. And Nichols himself does not deny that rules can be preserved from one generation to the next, even though they are *not* emotively laden. He writes: “*obviously* there are other important factors in cultural evolution. The hypothesis is only that affective resonance will be *one* of the factors that influence cultural evolution” (Nichols 2008, 270). It is clear, therefore, that the normative system cannot be explained exhaustively in terms of emotive causes: in particular, it is very plausible to suggest that some of our rules can be explained with reference to the aims that they serve, their favouring social cohesion, or their being supported by reasons that can be widely shared. In other words, lacking any proof of the fact that all our rules are based on an emotive sanction, it is likely that some (implicit or explicit) rational

mechanism exists that, along with our emotive reactions, helps to account for our normative theory. It is this supplement of rational reflection that allows to distinguish between the moral rules and other rules, such as those relative to disgust, that are supported by emotive mechanisms as well.

This shows that one of the factors explaining the adoption of the rules—even of those rules that are in fact supported by the emotive mechanism of concern—may well be the acknowledgement of objective reasons for their adoption, *e.g.* reasons consisting in the desire to protect people from unnecessary suffering and the like. In this picture, the emotive perception of others' suffering would merely favour and support the perception of these reasons: the reasons themselves, however, could not be reduced to the emotions, being rather their causes. The fact that we have the normative code that we have would thus partly be explained by a mechanism of reflective endorsement, in which several normative reasons—including those that are brought about by our automatic emotive responses—are reviewed in a process of rational deliberation. If this is so, the core moral judgments themselves are generated by emotions that have been shaped by the normative theory governing the process of education; and this normative theory carries the traces of a complex cultural process. Moreover, it is clear that the moral experience of adult individuals decisively depends on a normative theory that is no longer simply received, but self-consciously accepted and critically discussed. At this level, moral judgments may be partially independent from the actual activation of emotive responses, as in fact they generally are: it is a datum of experience that many of our ordinary moral judgment are not associated to, nor in any way generated by, the on-line activation of moral emotions.

The conclusion to be drawn is that, while Nichols' criticism seems to be effective against extreme rationalistic views, according to which reason provides a full explanation of morality, quite independently of emotive or sentimental reactions, it does not rule out more moderate rationalistic views. In fact, a) scientific evidence concerning the moral capacity of children and of psychopathic or "acquired sociopathic" individuals shows that some basic emotive capacity is a prerequisite of "the moral point of view", which is partly defined by the capacity to be emotionally tuned in with what happens to other people; and b) the normative theory that defines which emotions to have and which not, seems to be partly shaped by our emotive reactions. This, however, is compatible with a view according to which morality is, at root, the space of reasons, and in many cases (though not in all) moral judgments

depend on some rational (and fallible) deliberation on the reasons that we have. More precisely, in young children's core moral judgments the reference to reasons is implicit and indirect, since it is mediated by the normative theory, which, though represented internally, is not explicitly present to the consciousness of the judging individual. In adult individuals' more complex judgments, on the other hand, there is almost always some direct reference to our reasons. In fact, according to Nichols himself, an individual who lacked any normative theory would not make real moral judgments, but would limit herself to express her emotions, as if uttering an interjection. In other words, contrary to XXth century emotivists, Nichols' neosentimentalism does not treat emotions as in any way part of the content of moral judgments; emotions simply "play a role in leading us to treat as distinctive certain violations, including many of those we consider "moral", like violations of harming others" (Nichols 2008, 263). This, however, is quite compatible with a rationalistic view according to which emotive reactions are primary sources of reasons to act, and moral judgments are the reflective endorsement of our best reasons, mediated by the normative theory.

5. Let us now move to Nichols' thesis c), that is, the objection to motivational internalism: this is the view according to which there is a conceptual link between accepting a moral judgment and having some motivation to act accordingly. Nichols believes that empirical evidence on psychopaths shows the possibility to make moral judgments without acquiring the corresponding motivations in the least degree; however, it is far from obvious that the moral judgments of psychopaths should count as authentic. On this issue, Nichols is happy to rely on the empirical evidence of general opinion: he simply says that most people do believe that these moral judgments are authentic (Nichols 2002c e 2004). Even though we should accept this conclusion, it could hardly count as a decisive refutation of *rationalistic* internalism: according to this view, in fact, it is a conceptual truth concerning moral judgments that they provide reasons for action, which in turn provide motivations to act, *so far as we are rational* (Korsgaard 1986; Smith 2004 and 2008; Joyce 2008). And rationalistic internalism acknowledges the existence of many reasons that may restrict the motivational capacity of moral reasons: for one thing, weakness of the will, or the fact that, acknowledging the existence of a reason to do *x*, we acquire a motive to do it, but we are pulled by a stronger motive to do *y*; and other forms of practical irrationality, on account of which we fail to acknowledge the practical implications of what we sincerely declare to believe. Doubtless, psychopathy is itself a motive inhibiting the normal functioning of practical rationality; as long as they constitute blatant



examples of irrationality, psychopaths cannot constitute an objection to the view that to accept a moral judgment and not to have any corresponding motivation is to be irrational, even if we should accept that their moral judgments are authentic.

On the other hand, to affirm the authenticity of these judgments is scarcely plausible. The very fact, stressed by empirical research, that practical reason is not a cold calculation of means and ends, but a cooperative enterprise between emotive responses and reflective processes, shows that the moral judgments made by psychopathic individuals are not authentic. As a matter of fact, they do not adequately distinguish between conventional and moral norms and therefore, in contrast to three-year old children, do not possess the moral capacity, according to Nichols' criterion<sup>1</sup>: we might say that they but quote socially widespread moral beliefs and rules. These patients lack the mechanism of concern that helps understand the seriousness of moral violations and that motivates moral behaviour. In the standard terms of contemporary metaethical discussion, the moral judgments of psychopathic individuals are paradigmatic examples of an "inverted commas" use of moral language. In any case, the fact that psychopaths make real moral judgments might be an argument against "simple motivational internalism", according to which moral judgment *guarantees* moral motivation, but not against the form of internalism peculiar to moral rationalism (Joyce 2008).

Similar considerations can be offered with reference to patients with ventromedial lesions of the prefrontal cortex. Since these patients do make moral judgments, but lack any motivation to act correspondingly, they have been considered concrete examples of amorality and direct proofs of the falsity of motivational internalism. As noted by Adina Roskies, they are "walking counterexamples to this internalist thesis" (Roskies 2003, 51). Again, how authentic these moral judgments can be, is a matter of discussion: and if they are, the fact that rational and emotive areas of their brain are unrelated justifies the conclusion that VM patients' practical rationality is compromised; therefore, it is difficult to see how they could undermine the peculiar motivational internalism of the rationalistic view. On the other hand, it must also be stressed that there is a radical difference between making moral judgments in third and in first person: VM patients may be able to make "abstract" moral judgments from a spectator's point of view, but are definitely unable to make first person moral judgments,

---

1 Of course, one could deny that the capacity to distinguish between the moral and the conventional is constitutive of the moral faculty, as suggested by Roskies (2008).

since this would imply a commitment to act accordingly that they are unable to adopt. In this sense, although it may be incorrect to say that VM patients who are able to reason abstractly on moral dilemmas in Kohlberg's style make "inverted commas" moral judgments—for they seem seriously convinced of the truth of what they say—it is nonetheless incorrect to say that they make authentic moral judgments. To do this, they should have the capacity to decide in first person, that is, to assume the agent's perspective and to understand that action *x* is what is required *from me in these circumstances* (Kennett and Fine 2008).

6. One last word on the issue of moral objectivism (Nichols' thesis *d*), or the idea that the sentimental rule approach and its story concerning the cultural fitness of emotively laden rules undermine moral objectivity.

For one thing, it must be noted that moral rationalism is committed to moral objectivism, but not necessarily to moral realism: for example, a rationalist view such as Greene's or Singer's affirms the necessity to reformulate our evolutionary-biased normative theory in a consequentialist shape to reach moral objectivity, but does not embrace moral realism, nor justify any belief in objective moral *facts* (Singer 2005; Greene 2003; Greene 2008).

On the other hand, we cannot exclude that emotions might point out moral facts: they might not be contingent evolutionary facts, but means to acquire knowledge of some independently existing moral reality. Of course, the plausibility of this view depends on how such a "moral reality" is conceived: if to accept that there are moral facts is to say that, in each situation, there are objective moral reasons to act, to claim that the emotions we have are the "right" ones because they point out the moral facts is simply to claim that in the emotive reactions accompanying the consciousness of some other's suffering or being treated unjustly, we have access to objective reasons to avoid that suffering or injustice: reasons, that is, that can be acknowledged and shared by everyone. Rational reflection can detect these reasons and decide whether to adopt them, that is, whether to transform them in maxims of action and possibly in general principles of our normative theory. This does not alter the fact that emotive reactions are major sources of such objective reasons that our reflection can sanction.

In other words, the rationalistic view affirming the practical character of reason does not necessarily (nor generally) maintain that moral statements refer to purely objective facts, utterly independent of any emotion or

practical interest of the agent. It claims that they express the reasons for action that are highlighted both by our emotive reactions and by our rational reflection, while facing practical situations. Moral choice, therefore, expresses both our emotive and rational nature, namely, the embodied nature of our moral agency.

## REFERENCES

- Berker, S. (2009), "The Normative Insignificance of Neuroscience", *Philosophy and Public Affairs*, 37, 2009, pp. 293-329;
- Greene, J. (2003), "From Neural 'Is' to Moral 'Ought': What Are the Moral Implications of Neuroscientific Moral Psychology?", *Nature Reviews Neuroscience*, 4, 2003, pp. 847-850;
- Greene, J. (2008), "The Secret Joke of Kant's Soul", in W. Sinnott-Armstrong (ed.), *Moral Psychology. Volume 3: Emotion, Brain Disorders, and Development*, MIT Press, Cambridge, Mass., pp. 35-79, 2008;
- Greene, J. et al. (2001), "An fMRI Investigation of Emotional Engagement in Moral Judgment", *Science*, 293, 2001, pp. 2105-2108;
- Greene, J. et al. (2004), "The Neural Basis of Cognitive Conflict and Control in Moral Judgment", *Neuron*, 44, 2004, pp. 389-400;
- Greene, J. et al. (2008), "Cognitive Load Selectively Interferes with Utilitarian Moral Judgment", *Cognition*, 7, 2008, pp. 1144-1154;
- Haidt, J. (2001), "The Emotional Dog and its Rational Tail: A Social Intuitionist Approach to Moral Judgment", *Psychological Review*, 108, 2001, pp. 814-834;
- Haidt, J. and Joseph, C. (2004), "Intuitive Ethics: How Innately Prepared Intuitions Generate Culturally Variable Virtues", *Daedalus*, 133, 2004, pp. 55-66;
- Joyce, R. (2008), "What Neuroscience Can (and Cannot) Contribute to Metaethics", in W. Sinnott-Armstrong (ed.), *Moral Psychology. Volume 3: Emotion, Brain Disorders, and Development*, MIT Press, Cambridge, Mass., 2008, pp. 371-394;
- Kennett, J. and Fine, C. (2008), "Internalism and the Evidence from Psychopaths and 'Acquired Sociopaths'", in W. Sinnott-Armstrong (ed.), *Moral Psychology. Volume 3: Emotion, Brain Disorders, and Development*, MIT Press, Cambridge, Mass., 2008, pp. 173-190;
- Korsgaard, C. (1996), "Skepticism About Practical Reason", in *Creating the Kingdom of Ends*, Cambridge University Press, Cambridge, 1996, pp. 311-334;
- Nichols, S. (2002a), "Norms with Feeling: Towards a Psychological Account of Moral Judgment", *Cognition*, 84, 2002, pp. 221-236;
- Nichols, S. (2002b), "On the Genealogy of Norms: A Case for the Role of Emotion in Cultural Evolution", *Philosophy of Science*, 69, 2002, pp. 234-255;
- Nichols, S. (2002c), "Is It Irrational to Be Immoral? How Psychopaths Threaten Moral Rationalism", *The Monist*, 85, 2002, pp. 285-304;
- Nichols, S. (2004), *Sentimental Rules. On the Natural Foundations of Moral Judgment*, Oxford University Press, Oxford, 2004;

- Nichols, S. (2008), "Sentimentalism Naturalized", in W. Sinnott-Armstrong (ed.), *Moral Psychology. Volume 2: The Cognitive Science of Morality: Intuition and Diversity*, MIT Press, Cambridge, Mass., 2008, pp. 255-274;
- Roskies, A. (2003), "Are Ethical Judgments Intrinsically Motivational? Lessons from 'Acquired Sociopathy'", *Philosophical Psychology*, 16, 2003, pp. 51-66;
- Roskies, A. (2008), "Internalism and the Evidence from Pathology", in W. Sinnott-Armstrong (ed.), *Moral Psychology, Volume 3: The Neuroscience of Morality: Emotion, Brain Disorders, and Development*, MIT Press, Cambridge, Mass., 2008, pp. 191-206;
- Singer, P. (2005), "Ethics and Intuitions", *The Journal of Ethics*, 9, 2005, pp. 331-352;
- Smith, M. (2004), *The Moral Problem*, Basil Blackwell, Oxford, 2004;
- Smith, M. (2008), "The Truth About Internalism", in W. Sinnott-Armstrong (ed.), *Moral Psychology, Volume 3: The Neuroscience of Morality: Emotion, Brain Disorders, and Development*, MIT Press, Cambridge, Mass., 2008, pp. 207-215.