

The New Ethics of Neuroethics

TOM BULLER

Abstract: According to a familiar distinction, neuroethics incorporates the neuroscience of ethics and the ethics of neuroscience. Within neuroethics, these two parts have provoked distinct and separate lines of inquiry, and there has been little discussion of how the two parts overlap. In the present article, I try to draw a connection between these two parts by considering the implications that are raised for ethics by scientific findings about the way we make moral decisions. The main argument of the article is that although neuroscience is “stretching” ethics by revealing the empirical basis of our moral decisions and, thereby, challenging our present understanding of the dominant ethical theories, substantial further questions remain regarding the impact that neuroscience will have on ethics more broadly.

Keywords: moral theory; neuroscience; moral enhancement; ethics; neuroethics

Introduction

Over the past few years, a number of major national and international brain projects have been launched.¹ The establishment of these projects and their substantial funding gives credence to the view that the search to understand the brain is the dominant scientific inquiry of our time. As these projects move forward, an important question to ask is: What impact, if any, will our increasing knowledge of brain function have on our moral thinking? This question has already prompted considerable discussion in neuroethics, where it has been framed in terms of the purported challenge posed by neuroscience to our legal and moral notions of agency and responsibility. In this article, I will consider two lines of argument that suggest that neuroscience can advance, rather than challenge, morality: first, by revealing the basis of our moral intuitions and thereby revising our moral framework; and second, by enhancing our moral behavior through pharmacological (and other) means. One reason to think that this is an important question to address is that it examines neuroethics itself and, in particular, the relationship between the neuroscience of ethics and the ethics of neuroscience.

“Stretching Ethics to the Breaking Point”

In a well-known and influential article, Holmes Rolston III argued that environmental ethics “stretches classical ethics to the breaking point.”² “Classical ethics” (CE) is a framework with which we are broadly familiar: one in which the supposedly unique capacities of humans are regarded as having moral priority and there is a clear separation between the human and the natural; in other words, an ethic that privileges rationality and conscious decisionmaking, and that sees the moral agent as an impartial, rational observer disconnected from context and environment. According to Rolston, CE is too narrowly defined and is therefore unable to accommodate a variety of nonhuman objects of duty, such as animals, organisms, species, and ecosystems. This unwarranted exclusion is the result of CE privileging capacities that arguably are unique to humans—“singularity, centeredness, selfhood, and individuality”—over those that are common to both humans and other living entities. A different but related criticism focuses not on the exclusion

of nonhuman objects of duty, but on the privileging of rationality, impartiality, and individuality, which reflects a gendered bias. A credible ethics has to include in its moral framework care and compassion, and our relationships to others (both human and nonhuman). In this regard, ethics is political: it should seek as its goal the empowerment of oppressed and marginalized groups and individuals.^{3,4}

In different ways, the abovementioned criticisms challenge the way that morality and the moral agent is separated from the world; in brief, we might say that CE needs to be more empirically informed. One way that ethics can be more empirically informed is if its normative expectations and requirements are constrained by our moral psychology: the influences (internal and external) that shape the way that we make moral decisions. As Owen Flanagan says:

“The Principle of Minimal Psychological Realism. Make sure that when constructing a moral theory or projecting a moral ideal that the character, decision processing and behavior prescribed are possible, or are perceived to be possible, for creatures like us.”⁵

Presumably, the merit of adopting this principle is in part pragmatic: a moral theory or ideal that is impossible or perceived to be so will have limited success and few adherents. According to what may be termed the more “traditional” view, facts about our moral psychology have little, if anything, to do with our moral obligations. It may be an interesting descriptive and empirical fact that we are hardwired or enculturated to make certain types of moral judgments or to have certain types of moral intuitions, but this has no relevance to the normative question of what we *ought* to do. The traditional perspective thus adheres to the view that one cannot legitimately move from an “is” to an “ought,” that is to say, from descriptive to normative claims. Understood broadly, this position would appear to imply that the neuroscience of ethics and the ethics of neuroscience have little connection, because if we regard claims about brain function as descriptive and discussion of the ethics of neuroscience as inherently normative, then these two parts of neuroethics would appear to be essentially divided. It is precisely this aspect of the traditional view, however, that recent work in the neuroscience of ethics is challenging.

In a number of publications Joshua Greene has presented extensive evidence from neuroscience and other fields indicating that our current moral framework and our understanding of its dominant theories is inaccurate. This conclusion is important not only in and by itself, but also because it provides a significant example of the how the neuroscience of ethics can shape the ethics of neuroscience. According to Greene “Science can advance ethics by revealing the hidden inner working of our moral judgments, especially the ones that we make intuitively. Once these inner workings are revealed, we may have less confidence in some of our judgments and the ethical theories that are (explicitly or implicitly) based on them.”⁶

Before looking at the way in which science might lessen our confidence in our current ethical framework, it is helpful to briefly describe its dominant theories.

There has been considerable discussion in neuroethics on the use of neuroimaging to “read brains” to determine whether a person is lying, for example. In considering whether such a practice is morally defensible, it is reasonable to suppose that our deliberations will be informed by either consequentialist or deontological moral thinking. In very broad terms, consequentialism holds that actions are right

or wrong in terms of their consequences, and that these can be evaluated in a number of ways. For example, a historical influential position, utilitarianism, evaluates consequences in terms of the amount of overall happiness produced. In making a moral judgment, the agent is expected to follow the course of action that leads to the best overall consequences. In contrast, deontology holds that the agent's intention and reason for performing an action matter rather than do its consequences. Through rational deliberation, the moral agent identifies the course of action that is consistent with the autonomy of the agent and others. An action is morally praiseworthy if it is performed out of sense of moral obligation, rather than because of sentiment or for consequential reasons. A common (although controversial) perspective is that consequentialism is less dogmatic than deontology, because whereas the deontologist might decide that a course of action such as lying to help a friend, is "just wrong," the (act) consequentialist would maintain that the rightness or wrongness of the action is contingent on factors relating to the particular circumstances in which the action occurs.

According to Greene, science challenges the abovementioned account by providing evidence to suggest that "characteristically deontological judgments" are more emotional and less deliberative than "characteristically consequentialist" ones. Furthermore, deontological judgments tend to be automatic and associated with "up-close-and-personal" situations, and the moral judgments that are made reflect not rational deliberation but post facto rationalizations. In brief, (neuro)science indicates that consequentialist intuitions and judgments are more cognitive and rational than deontological ones.^{7,8,9,10}

To illustrate the matter, Greene considers the well-known *Switch* and *Footbridge* cases.

Switch. A runaway trolley is headed for five people who will be killed if it proceeds on its present course. The only way to save these people is to hit a switch that will turn the trolley onto a side track, where it will run over and kill one person instead of five.

Footbridge. A runaway trolley threatens to kill five people, but this time you are standing next to a large stranger on a footbridge spanning the tracks, in between the oncoming trolley and the five people. The only way to save the people is to push this stranger off the bridge and onto the tracks below.¹¹

In broad terms, most people judge it to be permissible to hit the switch in *Switch* but impermissible to push the stranger in *Footbridge*. According to Greene, this difference in our moral judgments reflects not rational deliberation or deep philosophical thinking, but the fact that "needy people who are up-close-and-personal push our moral buttons."¹² The difference can be explained by the fact that *Footbridge* necessitates a far greater degree of personal force and contact. As Peter Singer and others have argued, however, there are good reasons for thinking that although these elements matter psychologically, their moral relevance is questionable: I may be more aware of the impact of poverty in the United States than in the Sudan and care more about those close to home, but this does not mean that the hardship and suffering of those in the Sudan is any less, or less important, than of those in the United States.

The conclusion that Greene draws is that we have reason to “tilt towards consequentialism.” If the evidence suggests that our deontological judgments reflect intuitions that are based on questionably relevant factors and immediate emotional responses, and we continue to hold the view that ideal moral judgments should privilege rational deliberation, then consequentialism is the superior theory. Accordingly, a revised moral framework would privilege consequentialist over deontological moral thinking.

There have been a number of responses to the argument that science can advance ethics as described previously.¹³ One of these is to accept the scientific evidence but insist that the normativity of moral judgments cannot be determined by descriptive facts about our moral psychology and brain function. Therefore, even if we grant on empirical grounds that deontological judgments are more emotional and automatic than consequentialist ones, this leaves intact and unanswered the normative question of whether it is morally permissible to kill one to save five. This response may, however, be unsatisfactory, because it can be argued that it somewhat misses the main point; namely, that our answer to the normative question reflects scientific facts about our moral psychology: the deontological judgment, for example, that pushing the stranger off the footbridge is “just wrong” is based on the fact that the interaction in this instance is up-close-and-personal.

It is unlikely, however, that this reply will convince someone who insists on a clear distinction between the descriptive and the normative. If we agree with this position, then we are likely to hold that the two parts of neuroethics—the neuroscience of ethics and the ethics of neuroscience—are distinct, and that our evaluation of ethical issues in neuroscience should be independent of empirical evidence about the way that we make these evaluations.

Judgments, Principles, and Proximity

It is reasonable to suppose that the conclusion to “tilt toward consequentialism” would be resisted by those like Bernard Williams who are critical of this approach to ethics.¹⁴ According to Williams, consequentialism makes personal integrity more or less unintelligible, because the theory requires us to abandon our personal projects, and dictates that we are not especially responsible for what we do, as opposed to what someone else does. To understand Williams’s criticism, it is helpful to return to *Switch* and *Footbridge*. According to the revised moral framework, the “squeamishness” that one might feel at having to push the stranger off the bridge should be judged to be “irrational,” that is, not rational and of limited moral value. Nevertheless, although of limited value, the (moral) psychological fact of the matter is that typically this is how a person feels when faced with this type of decision, and that this feeling informs our moral intuitions. Presumably, if we adopt the revised moral framework in light of the scientific evidence, then when faced with *Footbridge*, we should put our squeamishness aside and accept that the up-close-and-personal nature of the interaction, and our own personal involvement with the situation, are morally irrelevant; in other words, we should adopt a robust agent-neutral point of view in making the appropriate moral judgment.

Furthermore, one can make the point more vivid if we suppose that the person faced with *Footbridge* realizes that the stranger is, in fact, a good friend or a close relation. According to the revised ethical framework, this fact carries no

moral weight, because it can be seen as just another form of proximity and, therefore, to be discarded. As Williams contends, however, this demand threatens to alienate people from their own feelings and values.

One option that a supporter of the revised framework might make in response to the abovementioned criticism is to maintain that personal feelings and values do not matter morally. It may well be the case that one feels more squeamishness at having to push one's best friend rather than a stranger off the footbridge, but it does not follow that the feeling (or its degree) is morally relevant. Furthermore, one should not frame the matter in terms of killing one's best friend *or* a stranger, for the choice is between saving five people by pushing a switch and thereby killing one's best friend, or saving five people by pushing one's best friend off the bridge. Therefore, even if the personal feelings of the agent do matter, they are not decisive.

In what may be a consequence of the obvious novelty of emerging neurotechnologies, there is little discussion in neuroethics of "neuroethical dilemmas." Instead, the focus has been on broader questions; for example, the ethics of cognitive enhancement or the development of guidelines for the responsible conduct of neuroscientific research. Nevertheless, in light of the foregoing, we can ask how the revised moral framework might inform specific cases. Consider the case of a university administrative committee tasked with determining whether students should be prohibited from using Ritalin (methylphenidate) to enhance cognition. It is likely that in their deliberations, the committee will think of particular cases as means to reveal and clarify their intuitions, and we can imagine that the intuitions of those of a more old-fashioned deontological persuasion will be that the use of Ritalin for this purpose is "just wrong" and a clear case of cheating. Differently, we might imagine another committee tasked with formulating principles for the ethical use of nonhuman primates in neuroscience research. In this case, we can imagine that there will be those who believe this practice to be justified on the grounds that the use of nonhuman primates is essential for the development of treatments, and, in opposition, those who reject such consequentialist arguments on the grounds that nonhuman primates have moral status and deserve (greater) protection.

We might ask what role and weight should be given to the deontological judgments and opinions in these two cases. If we adopt the revised moral framework, then we might say that in their deliberations, the committees should take into account the limitations of deontological judgments that result from the more emotional intuitions on which they are based. If the goal is to make the most rational decision, then we should privilege consequentialist judgments. One objection to this position is that it is by no means unanimously agreed that we should give pride of place to reason. A further objection is that even if it is granted that consequentialist judgments are more rational, we would need an additional line of argument to show that we ought, therefore, to privilege such judgments.

The matter becomes more complex if we consider the relationship between cases and principles in the broader moral theory. For the sake of argument, we can focus on the principle, "Do unto others as you would have them do unto you." This principle can be part of both consequentialist and deontological perspectives; a Rule Utilitarian, for example, could advocate the rule on the grounds that adherence to it increases net utility. The generation and articulation of the rule has little to do with the psychology of the agent in specific cases, because as the principle is

necessarily general, it cannot be contingent on the circumstances of particular cases. Therefore, although we have an empirical explanation for the difference in judgments that people make when faced with *Switch* and *Trolley*, and we may have reason to privilege one type of moral reasoning over the other, the moral rules and principles that play a part in the moral decisionmaking process are independent of the empirical evidence. (As Flanagan contends, our moral rules should be psychologically realistic if we wish them to have the best chance of being respected, but this speaks more to the standards that we set and to our expectations, rather than to the generation of the rules themselves).

Moral Enhancement

One topic in the neuroscience of ethics that would seem to be particularly relevant to the present discussion is moral enhancement. According to an influential definition, moral enhancement can be understood as the attenuation of countermoral emotions.¹⁵ As Tom Douglas says “A person morally enhances herself if she alters herself in a way that may reasonably be expected to result in her having morally better future motives, taken in sum, than she would otherwise have had.”¹⁶

Similarly, Julian Savulescu and Ingmar Persson claim, “To be morally enhanced is to have those dispositions which make it more likely that you will arrive at the correct judgment of what it is right to do and more likely to act on that judgment.”¹⁷

In support of his argument, Douglas presents the following case: “[*The Biased Judge*] James is a district court judge in a multi-ethnic area. He was brought up in a racist environment and is aware that emotional responses introduced during his childhood still have a biasing influence on his moral and legal thinking. For example, they make him more inclined to counsel jurors in a way that suggests a guilty verdict, or to recommend harsher sentencing, when the defendant is African-American. A drug is available that would help to mitigate this bias.”¹⁸

Although Douglas is careful to point out that his account of moral enhancement does not endorse any one type of moral theory or position, the “attenuation of counter-moral emotions” would seem to have more resonance with consequentialism than deontology, because according to a traditional Kantian view of deontology, only the will is “good in itself,” and emotion is neither moral nor countermoral. Accordingly, the attenuation and subsequent changes in disposition will not, *prima facie*, lead to moral enhancement, because emotions and their consequences have little role in moral action. Consistent with this conclusion, Savulescu and Persson present the following as an example of an “enhanced utilitarian”

- 1) Cognitive enhancement—to accurately estimate the consequences of action and the impact on people’s preferences
- 2) Impulse control to enable one to act on one’s judgements of right action
- 3) Willingness to sacrifice one’s own preference satisfaction for the satisfaction of other’s preference.¹⁹

This notion of moral enhancement would appear to fit very well with the revised ethical framework: people morally enhance themselves through attenuating countermoral emotions, sacrificing personal preferences, and making their judgments more deliberative; in other words, by changing moral psychology in a way that

echoes the reasons that make consequentialist moral judgments superior to deontological ones. In this regard, Douglas' *Biased Judge* is similar to the person who is reluctant to push the stranger off the footbridge, because both are impeded from doing the right thing by their countermoral emotions: the Judge's moral judgments are biased against African-Americans as a result of the emotional responses introduced during his childhood; the moral judgment of the agent in *Footbridge* is "biased" as a result of the emotional valence of the up-close-and-personal situation.

Despite their initial similarity, however, the cases of *Biased Judge* and *Footbridge* are different in a key way, and this difference has important implications for the conclusion that science can advance ethics. This difference is illustrated by a revised version of the *Biased Judge* case, the *Socially Conscious Judge*, "[The *Socially Conscious Judge*] Joan is a district court judge in a multi-ethnic area. She was brought up in a racist environment and is aware that emotional responses introduced during her childhood still have a biasing influence on her moral and legal thinking. For example, they make her more inclined to counsel jurors in a way that suggests a not-guilty verdict, or to recommend more lenient sentencing, when the defendant is African-American. A drug is available that would help to mitigate this bias."²⁰

On the basis of the supposed superior merits of consequentialist moral judgments and rational deliberation, it would seem that Joan should take the drug to mitigate the bias. Moreover, there is reason to think that this would be an appropriate use of moral enhancement as described by Savulescu and Persson. What is less clear, however, is whether we should regard Joan's emotions as "countermoral," or if the attenuation of these emotions can be reasonably expected to result in morally better future motives than she had before. For we might think that it is entirely appropriate for Joan to be "biased" in this way and for this "bias" to have good consequences, for example, by reducing the discrimination against African-Americans in the legal system. This point highlights the difficulty in determining what counts as a "countermoral emotion," and the role that emotion plays in moral enhancement and in the revised framework. On the one hand it can be argued that Joan *ought* to take the drug to reduce her bias toward African-American defendants because it is emotional, personal, and non-deliberative, and will lead to *bad* consequences—justice, after all, is meant to be blind; on the other hand it can be argued that she *ought not* to take the drug precisely because the emotion is not counter but "pro-moral"—the sentiments that produce it are laudable (fairness and justice) and society is benefited by the motives of its members being influenced in this way and the consequences that result.

Conclusion

In thinking about the way that science can advance ethics, we very quickly come up against nonscientific questions that cannot be answered empirically. This should not be understood as a criticism or rejection of the view that science can advance ethics, because as the case of moral enhancement indicates, it may well be true that by taking a particular drug, a person can make moral judgments that are more deliberative and that lead to significantly better consequences. The challenge, however, is that we would still need to address the question of whether altering our judgments makes them more moral and enhances the person's morality.

Environmental ethics and feminist ethics stretch CE by revealing the bias in classical ethics and by insisting that ethics be situated within the broader environmental, relational, and sociopolitical context. Neuroethics has the potential to stretch classical ethics in a different way by challenging the perceived autonomy of moral judgment and the traditional view of the moral agent. As we move forward, neuroethics itself will need to consider the implications of the neuroscience of ethics and the extent to which neuroscientific findings should be incorporated into neuroethical analysis. It may be the case that the prevailing view will be that empirical findings have only a very limited impact on the normative aspect of the field. If this view does prevail, then a further important question to ask is what implications does this division between the ethics of neuroscience and the neuroscience of ethics have for the field of neuroethics.

Notes

1. Australian Brain Alliance, BRAIN Initiative, Human Brain Project, Canada Brain Research Fund, China Brain Project, Cuban Human Brain Mapping Project (CHBMP), Israel Brain Technologies, Latin American Brain Mapping Network (LABMAN), Brain Mapping by Integrated Neurotechnologies for Disease Studies (Brain/MINDS), Korean Brain Initiative, and Blue Brain Project.
2. Rolston H III. Environmental ethics: Values in and duties to the natural world. In: Bormann FH, Kellert ST, eds. *The Broken Circle: Ecology, Economics and Ethics*. New Haven: Yale University Press; 1991:228–47.
3. Merchant C. Ecofeminism and feminist theory. In: Diamond I, Orenstein G, eds. *Reweaving the World: The Emergence of Ecofeminism*. San Francisco: Sierra Club Books; 1990:77–83.
4. Warren KJ. The power and promise of ecological feminism. *Environmental Ethics* 1990:125–45.
5. Flanagan O. *Varieties of Moral Personality*. Cambridge, MA: Harvard University Press; 1993, at 32.
6. Green JD. Beyond point-and-shoot morality: Why cognitive (neuro)science matter for ethics. In: Liao SM, ed. *Moral Brains: The Neuroscience of Morality*. New York: Oxford University Press; 2016, at 119.
7. Greene JD. The secret joke of Kant's soul. In: Sinnott-Armstrong W, ed. *Moral Psychology*, Volume III. Cambridge, MA: MIT Press; 2008:35–81.
8. See note 6, Greene 2016.
9. Greene JD. From neural 'is' to moral 'ought': What are the implications of neuroscientific moral psychology? *Nature Reviews Neuroscience* 2003;4:847–50.
10. Haidt J. The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review* 2001;8(4):814–34.
11. See note 7, Greene 2008, at 41–2.
12. See note 9, Greene 2003, at 849.
13. Berker S. The normative insignificance of neuroscience. *Philosophy and Public Affairs* 2009;37(4): 293–329.
14. Smart JJC, Williams B. *Utilitarianism For and Against*. Cambridge: Cambridge University Press; 1973.
15. Douglas T. Moral enhancement. *Journal of Applied Philosophy* 2008;25(3):228–45.
16. See note 15, Douglas 2008, at 229.
17. Savulescu J, Persson I. Moral enhancement, freedom, and the God machine. *The Monist* 2012; 95(3):399–421, at 406.
18. Douglas T. Moral enhancement via direct emotion modulation: A reply to John Harris. *Bioethics* 2013;160–8, at 161.
19. See note 17, Savulescu, Persson 2012, at 406.
20. See note 18, Douglas 2013, at 161.