



**Target Article**

# Neuroethics: A New Way of Doing Ethics

**Neil Levy**, Florey Neuroscience Institutes and Oxford Centre for Neuroethics

The aim of this article is to argue, by example, for neuroethics as a new way of doing ethics. Rather than simply giving us a new subject matter—the ethical issues arising from neuroscience—to attend to, neuroethics offers us the opportunity to refine the tools we use. Ethicists often need to appeal to the intuitions provoked by consideration of cases to evaluate the permissibility of types of actions; data from the sciences of the mind give us reason to believe that some of these intuitions are less reliable than others. I focus on the doctrine of double effect to illustrate my case, arguing that experimental results suggest that appeal to it might be question-begging. The doctrine of double effect is supposed to show that there is a moral difference between effects that are brought about intentionally and those that are merely foreseen; I argue that the data suggest that we regard some effects as merely foreseen only because we regard bringing them about as permissible. Appeal to the doctrine of double effect therefore cannot establish that there are such moral differences.

**Keywords:** morality/ethics, neuroethics, philosophy

Neuroethics is not just another branch of applied ethics. There are several reasons why this is true. For one thing, its scope is broader, encompassing not only concerns about the permissibility or advisability of using certain technologies to “read” minds, enhance capacities, or control behavior (all questions that are closely analogous to those pursued in bioethics) but also questions about what it means to be human, whether we have free will, the nature of knowledge and of self-knowledge: questions that are more traditionally the terrain of philosophy broadly conceived, rather than of applied ethics. In this paper, I focus on a second difference between other branches of applied ethics and neuroethics. Neuroethics alone, I argue, offers us the opportunity to learn about, refine, and even dramatically to alter the tools we use as applied ethicists.

Central to the applied ethicist’s toolkit is the generation of intuitions: unreflective responses to actual and imaginary cases. Our intuitions play a pivotal role, both in the construction of our theories, and in assessing the major questions with which we deal as ethicists. Neuroethics, I show, can shed light on the processes that lead to the generation of our intuitions; moreover, it can show that some of our intuitions are likely to be generated in ways that render them unreliable. Neuroethics can therefore allow us to sort through our intuitions, separating those that ought to be retained as conducive to truth from those that ought to be rejected as misleading. If it is true that intuitions are central to applied ethics, it can therefore allow us to make better moral judgments.

## A NEUROETHICAL EXAMINATION OF INTUITIONS

Roskies (2002) distinguishes two branches of neuroethics, which she refers to as the ethics of neuroscience and the

neuroscience of ethics. The ethics of neuroscience is neuroethics as applied ethics; it consists of ethical reflection on neuroscience, its practice, and the technologies to which it gives rise. The neuroscience of ethics is instead concerned with what the sciences of the mind can tell us about the nature of morality and morally relevant topics in philosophy. In its guise as the neuroscience of ethics, one of the topics of neuroethics is how intuitions are generated. What brain regions are involved and how do they function to bring the agent to think that a particular action is forbidden, permissible, or obligatory? One of the most interesting findings produced by this work is that not all intuitions are generated in precisely the same manner. Instead, different processes are involved in generating different intuitions. This opens up the possibility that intuitions that are on a par phenomenologically might differ in ways that are relevant to their justification. Some intuitions might be generated in response to good evidence, while others might be generated by irrational processes. If we can show this, we can begin to distinguish between our intuitions, regarding only some of them as possessing weight in the process of attempting to reach reflective equilibrium. Just as we reject certain perceptual seemings because we have well-justified theories which lead us to discount them (for instance, we do not take the fact that a stick looks bent in water as a reason to conclude that it is bent), so we might be led to reject certain moral seemings on the grounds that they are generated by morally irrelevant processes.

There is plentiful evidence that moral judgments can be generated by morally irrelevant processes. Consider here Jonathan Haidt’s model of moral reasoning (Haidt et al. 1993; Haidt 2001). Haidt argues that moral judgments are generated in ways that bypass conscious reflection. On

Address correspondence to Neil Levy, Florey Neuroscience Institutes, 161 Barry St., Carlton South, 3053 Australia. E-mail: neil.levy@philosophy.ox.ac.uk

contemplating an action, we experience a feeling—what I have been calling an intuition—and we go on to form a correlative judgment. That is, if the feeling is one of unease or disgust, we will judge that the action is impermissible, whereas if the feeling is one with a positive content, we will judge that the act is permissible or required. For Haidt, the role of reasoning is merely to defend these judgments, like a lawyer defending a client, rather than to produce the moral judgments. Elsewhere (Levy 2006), I have argued that Haidt underplays the role of reasoning in generating moral judgments: Though we may typically form moral judgments in the way he suggests, nevertheless those judgments might be responses that we have only because social norms have altered under the pressure of reasoning. The average American might not feel a strong disgust response toward homosexual acts, for instance, because social norms have changed, so that this response is no longer enculturated, and social norms have changed (in part) due to people becoming convinced that arguments for the wrongness of homosexuality are weak. However, this question is irrelevant here; we can utilize Haidt's model without settling it.

Haidt's model predicts that we will tend to form moral judgments in response to intuitions, no matter how the intuitions are generated. This suggests that subjects' intuitions can be manipulated. If we can generate feelings of unease or disgust, for instance, in subjects, we can bring them to make correlative moral judgments, even when the feeling has been generated by a morally irrelevant stimulus. This is not mere speculation; together with collaborators, Haidt has tested this proposal. Wheatley and Haidt (2005) used posthypnotic suggestion to generate a disgust response to the word "often." As a result, subjects who heard the word would experience a pang of disgust but, in the absence of some alternative explanation of the feeling, would regard it as a response to the situation that was being described. Subjects who had undergone this experimental manipulation judged scenarios of moral wrongdoing as more seriously wrong than did control subjects who got the same scenarios. Indeed, a large minority of subjects in the posthypnotic suggestion group found moral wrongdoing in situations entirely devoid of anything remotely wrong. The irrelevantly generated feeling caused a moral judgment. Schnall, Haidt, Clore, and Jordan (2008) used a similar paradigm to intensify judgments of moral wrongness. They found that for subjects who score in the upper half of a scale measuring consciousness of one's own body, being seated at a dirty desk led to stronger moral judgments.

What these studies show is that we can, in principle, begin to distinguish good intuitions from bad, but that we cannot do so from within. That is, we cannot begin to distinguish which of our intuitions are reliable simply by reflecting on them from the first-personal perspective. Instead, we need to adopt the third-personal perspective of science, examining how the intuitions were generated.

## ASSESSING INTUITIONS

In this section, I describe two sets of evidence for the claim that some intuitions are less reliable than others. The first, which uses neuroimaging evidence, is relatively well-known and I describe it only briefly. The second, using psychological evidence, is less well-known; I therefore describe it at more length.

### Deontological and Consequentialist Intuitions

The better known evidence comes from a neuroimaging study. Greene and colleagues (2001) used functional magnetic resonance imaging to examine the brains of subjects responding to moral dilemmas. The experimenters used probes modeled on the famous trolley problem (Foot 1978). In the trolley problem, we are forced to decide whether to perform an action that will result in the death of one person, who would otherwise survive, in order to save the lives of five. The dilemma results from the fact that subjects respond very differently to cases that share the same broad outlines—that is, in which the choice is between acting so that one person dies or so that five people die. Consequentialism counsels that we always choose the action that saves the five, but ordinary people, and most philosophers, do not always choose the option that consequentialism recommends. When the choice is between allowing the trolley to hit the five or diverting it onto a sidetrack on which it will hit and kill one, most philosophers and ordinary people judge that we ought to divert the trolley. But when the five can be saved only by pushing one person into the path of the trolley, most judge that the action is impermissible.

Marc Hauser and colleagues have amassed a huge database of responses to the trolley problem and related dilemmas, using a between-subjects design (that is, subjects do not view both members of a pair of dilemmas within a single session, since they might notice an apparent inconsistency in their responses and override their intuitions to impose greater consistency). About 90% of subjects judge it permissible or obligatory to redirect a threat to save five lives at the expense of one, but about the same number judge it impermissible to place the one between the five and the threat, though doing so would save the five at the expense of the one. Hauser and colleagues have attempted, on the basis of subjects' responses to these and related dilemmas, as well as by reference to subjects' justifications of their responses, to articulate the implicit principles that guide moral judgment (Hauser 2006; Cushman et al. 2006). They argue that some of these principles are cognitively accessible to subjects, whereas others—such as the principle that causing a harm by direct contact is worse than causing a harm at a distance—are neither easily accessible nor endorsed when subjects become aware of them.

Greene's study, however, has been interpreted not merely as articulating the principles guiding moral reasoning, but as providing evidence with regard to their rationality. Briefly, his team found that when subjects generated the responses endorsed by consequentialism, they exhibited

activation in regions of the brain associated with working memory. But when they gave responses that were more aligned with deontological judgments, regions associated with emotion showed significant activity, while those associated with working memory showed a degree of activation that was actually below the resting baseline. One way to interpret this results is as follows: Dilemmas with certain features—e.g., those in which better consequences are achieved by what is intuitively a causing, rather than allowing, of a harm—arouse a great deal of emotion in us, and this emotional arousal “crowds out” genuine reasoning. On this interpretation, deontological responses are irrational responses, the product of the fact that we cannot reason well under certain circumstances (Greene 2003; Singer 2005).

If Greene’s results stand up to scrutiny, they provide us with a strong reason to distinguish between different kinds of intuitions. Intuitions that are indistinguishable from a first-personal point of view, and that, from that perspective, we are justified as regarding as data for moral judgment formation, can be seen from a third-personal point of view to have radically different degrees of reliability. Since only our consequentialist intuitions are the product of reasoning, while our deontological intuitions are generated by some kind of morally irrelevant process, we ought to ignore the latter and side with the former. Of course, this finding would be directly relevant to the ethics of neuroscience. Since many of the standard arguments against, say, the use of cognitive enhancers are non-consequentialist—turning, for instance, on worries about the authenticity of the individual—we would have a strong reason to dismiss these arguments.

There is, of course, much more to be said about Greene’s findings. There are grounds, both philosophical and methodological, to question their validity. Some thinkers have challenged the probes used in this experiment, arguing that the cases used were not of the right kind to generate deontological and consequentialist intuitions (Kahane and Shackel 2008). Moreover, even if these methodological problem can be dealt with, it might be a mistake (the same mistake Haidt makes, in my view) to argue that a response is the product of reasoning only if the agent engages in reasoning at the time she generates the response. A judgment might be the product of reasoning as a community-wide enterprise, and an agent who makes the judgment therefore rationally justified in making it, even though that agent is unable to communicate, or perhaps even understand, the reasoning that has led to their judgment. In the language of epistemology, we can say that warrant can be transferred by testimony from a community of experts to laypeople (Coady 1992). Thus, for instance, when I say “Pluto isn’t a planet”, I may say something that is true, justified, and the product of reasoning, even though I may not know the reasons why Pluto isn’t a planet. Showing that our deontological intuitions are irrational and unjustified, similarly, requires showing more than that they are typically generated by a process that bypasses reasoning; it requires, further, showing that subjects do not

have these intuitions as a result of the reasoning of other agents.

### The Doctrine of Double Effect

Let me turn to the second piece of evidence that some intuitions are less reliable than others: evidence that seems less vulnerable to the criticisms just mentioned. This evidence comes from psychological studies of subjects’ responses, rather than from neuroimaging. Before I present the evidence, some background is necessary so that we can appreciate the moral relevance of the studies. The evidence I present suggests, I claim, that the intuitions that underlie the doctrine of double effect—a principle often invoked in moral reasoning—are in fact generated in ways that render them unreliable.

The doctrine of double effect was originally introduced and defended by Thomas Aquinas in the 13th century, in his *Summa Theologica*. In discussing self-defense, Aquinas points out that an action can have effects beside those that are intended. This fact, Aquinas argued, is morally crucial: In some circumstances it is permissible to cause effects when those effects are side effects of some intended goal, even though it would not be permissible to bring about those effects intentionally. For Aquinas, this principle justifies killing in self-defense: Though it is wrong to take a life, and therefore any intentional killing is impermissible, it is permissible to defend oneself, and if the only actions one can take to defend oneself bring about the death of an aggressor, then it is permissible to engage in that action. That is, killing in self-defense is permissible as long as the intention is not to kill the aggressor but instead to defend oneself.

Since Aquinas’ time, a great deal of effort has gone into clarification and development of the doctrine of double effect. Proponents of the doctrine now generally agree that it renders some otherwise prohibited actions permissible only when a number of background conditions are satisfied: The goal—the intended effect—must itself be morally good, the means to that good must be permissible, and the good must outweigh the unintended, merely foreseen, effect. In addition, there must be no other permissible means of bringing about the goal. When the appropriate conditions are satisfied, the doctrine of double effect justifies us in pursuing good goals even at the cost of bringing about morally bad states of affairs, as long as these states of affairs are merely foreseen and not intended.

Aquinas’ doctrine was immensely influential, and continues to play a central role in normative ethics today. Just war theorists appeal to it in order to justify certain actions resulting in the death of civilians; medical ethicists appeal to it to justify measures that hasten the deaths of terminally ill patients. The doctrine also plays a role in some neuroethical issues. For instance, opponents of cognitive enhancement might hold that there is a significant moral difference between taking a psychopharmaceutical intending to achieve some end other than enhancement, and intentionally enhancing oneself. The first might be permissible while the

second is impermissible, even though the effects on the agent might be the same. In all these cases, and many others with a parallel structure, causing an effect that it would (by deontological lights) be wrong to bring about intentionally is permissible, as long as it is a merely foreseen effect of an action intended to achieve a morally laudatory goal, and there are no other means to achieve that goal.

The doctrine of double effect has many critics. Some of the critics are consequentialists, who hold that we ought simply to aim to maximize the good, and not worry about which goals are intended and which mere side effects. Consequentialists deny the moral relevance of the distinction that the doctrine aims to preserve. It must be recognized, however, that the intuitions to which the doctrine appeals are, for many people, strong and tenacious. Unless we have some reason to regard them as unjustified—perhaps because generated by processes that are not responses to morally relevant information—we have good reason to regard these intuitions as data for moral theorizing.

In what follows, I argue that these intuitions are sensitive to moral considerations in a way that makes appeal to them question-begging. It is question-begging because agents' preexisting moral views influence the application of the doctrine in such a manner that it generates the appropriate output. Thus, I next claim, the doctrine of double effect cannot play any independent role in justifying the permissibility of certain kinds of actions. If I am successful, I will have produced a novel case for the claim that the doctrine is more rationalization than rational argument. I argue that the crucial distinction is normatively loaded—that is, dependent upon agents' prior moral views—and that invoking the doctrine is therefore circular: It is because we have the intuition that a certain action is permissible that we have the intuition that it is not intended. If that's right, invoking these intuitions in support of our moral claims is straightforwardly question-begging; it assumes much of what it aims to demonstrate. I don't say that the distinctions between intended and unintended consequences can never be drawn in a noncontroversial way; my claim is rather that the distinctions are too normatively loaded to serve as a criterion to sort difficult and controversial cases.

There is an obvious difficulty with the doctrine of double effect: It is far from obvious what an "intention" is. Some philosophers are Humeans about psychological states: They believe that all such states can be analyzed in terms of beliefs and desires. For a Humean, an intention is a certain combination of beliefs and desires (perhaps an intentional action is simply one that the agent believes will produce a state of affairs, coupled with a desire that that state of affairs be actual). This kind of conception of intention will distinguish between foreseen and intended actions on the basis, presumably, of the agent's desires: If an agent desires that the action bring about a certain state of affairs, then the agent intends it. Other philosophers depart from belief/desire psychology in holding that an intention is a distinct mental state, not reducible to combinations of belief and desires. For these philosophers, an agent who desires

that a state of affairs be actual might nevertheless fail to intend it, even as the agent performs an action which the agent believes will cause it to be actual. On this view, for instance, I might turn down an invitation to a party, foreseeing that by doing so I will cause offense and also while desiring to offend, without actually intending to offend. So these philosophers are obviously tasked with explaining how one can desire and cause X to occur without intending it to occur.

It is important to note that insofar as the doctrine of double effect is concerned, however, it doesn't matter whether the Humean or non-Humean account is the correct one. The doctrine distinguishes cases on the basis of how we attribute intentions to agents, not on how those agents actually are. To be sure, we can imagine a future, reconstructed, doctrine of double effect that sorts cases on the basis of the neural signatures of intentions, as detected by functional magnetic resonance imaging (fMRI) or electroencephalograph (EEG). But for the moment the doctrine of double effect asks us to judge the permissibility of actions by reference not to the actual mental states of agents, but to the kinds of mental states they are supposed, by the judge, to have. We need not know what intentions actually consist in, in order to attribute them to agents; it is even conceivable that we have a radically mistaken view of the nature of intentions, yet are reliable intention attributors. Indeed, since the doctrine is usually defended by reference to imaginary cases, it must be how we attribute intentions, and not the intentions agents actually have, that is directly at issue.

The central worry I sketch here concerns our mechanism of attributing intentions. According to the doctrine of double effect, an action is permissible if bad side effects are foreseen but not intended (and the other conditions are satisfied). According to the rival view I now sketch, a state of affairs that is a foreseen effect of an action that is (plausibly) held to aim at some other goal is judged to be unintended if (inter alia) the action is judged to be permissible. If that's right, then the doctrine of double effect will simply reflect the moral intuitions of its proponents; the rationale offered will be mere confabulation. That is, the permissibility judgment will not be an output of the doctrine; instead, the doctrine will generate a permissibility judgment only because of a prior assessment of the acceptability of the action.

Why do I suspect that the judgment about intention is not doing any work in motivating people's judgments about permissibility, but is in fact downstream of judgments of permissibility? Because there is now a great deal of work detailing the way in which intentionality judgments are sensitive to moral considerations. This phenomenon has come to be known as the Knobe effect, after Joshua Knobe, who discovered it (Knobe 2003, 2006). Knobe used a between-subjects design to discover what features of cases subjects respond to in judging whether a certain effect was intended or not. His evidence suggests that agents' prior moral views powerfully influence their attributions of intention.

Consider the following vignette:

The vice-president of a company went to the chairman of the board and said, "We are thinking of starting a new program. We are sure that it will help us increase profits, and it will also harm the environment." The chairman of the board answered, "I don't care at all about harming the environment. I just want to make as much profit as I can. Let's start the new program." They started the new program. Sure enough, the environment was harmed.

Knobe gave this vignette to subjects and asked them the following question: Did the chairman harm the environment intentionally? Most subjects (82%) said yes: The chairman intentionally harmed the environment (Knobe 2006).

But other subjects got this version of the story:

The vice-president of a company went to the chairman of the board and said, "We are thinking of starting a new program. We are sure that it will help us increase profits, and it will also help the environment." The chairman of the board answered, "I don't care at all about helping the environment. I just want to make as much profit as I can. Let's start the new program." They started the new program. Sure enough, the environment was helped.

Now, this is precisely the same story with the moral valence of the side effect changed: In the first version the chairman is deciding on a course of action that harmed the environment, whereas in the second the course of action will have the side effect of helping the environment. In both, we know a lot about the mental states of the actor: We know he intends to increase profits and doesn't care about the environment. But surprisingly, altering the moral valence of the side effect dramatically alters subjects' perception of its intentionality: The majority of subjects now judged that helping the environment was unintentional. Just 23% held that the effect was intentional (Knobe 2006). This finding gives us a *prima facie* reason to be suspicious of the justification offered in double effect cases. If judgments of intentionality are sensitive to moral considerations, then it might be because people judge the intentionality of a side effect on the basis of its moral permissibility, rather than judging the permissibility of an action on the basis of the intentionality (or unintentionality) of the side effect.

It might be objected that it doesn't matter which story is correct, whether subjects judge that certain kinds of harms are permissible because they judge them to be unintentional, or whether they judge these harms unintentional because they judge them to be permissible. The objection can be strengthened by drawing upon the popular linguistic analogy. Many thinkers now believe that moral competence should be understood on the model of Chomskyan linguistic competence (Mikhail, forthcoming; Hauser 2006). On this model, competence depends on a set of implicit rules that are innate; cultural and individual variation are explained by reference to parametric switches which may be set in one of several ways. Suppose this model, or something like it, is true of morality. In that case, the doctrine of double effect might simply be a mistaken theory about some aspect of our genuine moral competence. It would be akin to a mistaken grammatical theory, say, about why verbs must agree

with their subjects. Though the theory is mistaken, it would still be the case that verbs must agree with their subjects. Competence is one thing; theories about competence are another. So we might have a mistaken theory about an action's permissibility due to unreliable judgments about intentionality; this doesn't alter the fact that our moral competence allows us to sort cases into permissible and impermissible harms, nor does it alter the fact that the doctrine might yield the right answer when it is applied to sort cases into the permissible and the impermissible.

This is a forceful objection. As long as we are confident of our intuitions, if our explanatory theory regarding them fails we should search for a better theory, not abandon our intuitions. It is because we are confident in our linguistic intuitions—it is uncontroversial that subjects and verbs should agree—that the failure of the relevant theory should not trouble us with regard to these intuitions. But this is not in fact the state of play with the doctrine of double effect: The controversy with regard to the doctrine of double effect concerns how it sorts cases into permissible and impermissible, not the best way of capturing uncontroversial intuitions that we all share. Defenders of the doctrine hope to use the theory to sort cases that are controversial: They point to cases upon which agreement about permissibility is widespread, and say that because such cases have feature X (where X is the fact that a harm is foreseen but not intended, plus the other conditions mentioned earlier), any case that also has X should likewise be judged to be permissible. The theory is supposed to do some further work, not merely describe an existing competence. The intuitions to which the proponents appeal are supposed to justify a principle—that foreseen but not intended effects have a different moral status than those that were intended—that will appear among those we accept when we reach reflective equilibrium. So if it is the case that the principle of double effect is mistaken, we cannot be sanguine.

Under the rival theory I have sketched, we sort cases like "the program" discussed in the Knobe examples into the permissible and the impermissible by reference to irreducibly normative intuitions. Might we not invoke this revised theory to sort cases—in other words, doesn't this new theory identify a new feature X (i.e., our intuitions about the permissibility of an effect) with which to sort cases? Apparently not; the feature X by which cases are sorted seems to be agents' preexisting moral judgments about the effect of the case or action itself, and it is these moral judgments that are in dispute. If we sort cases by reference to intuitions, and these intuitions differ between subjects, then we cannot sort those cases in an objective way by reference to factor X. To put it another way, proponents of the doctrine of double effect appeal to the principle to say why we should accept their intuitions and not those of their consequentialist rivals; without the doctrine to back them up, they have nothing to appeal to in order to break the deadlock. Indeed, plausibly matters are even worse for them than this suggests: Since the consequentialists do have a theory to appeal to in order to justify their intuitions, and their theory is one with a great deal of intuitive appeal, if the doctrine of double effect

fails we will therefore have some reason to side with the consequentialists.

Mightn't it be objected that there is sufficient agreement on cases to isolate factor X? If the doctrine of double effect sorts some cases in the ways that its proponents urge, and the best explanation for this sorting turns on the intentionality of the relevant effects, then this is sufficient to justify the doctrine. There is, in fact, little doubt that there are many cases in which we will get a very high degree of agreement across subjects with regard to whether a particular foreseen effect is intended or not, and little doubt that on some of these cases there will be substantial agreement regarding moral permissibility. But that's precisely what we ought to expect, if our intuitions are responsive to moral considerations other than those to which the doctrine refers. Since that's the case, the proponent of the doctrine owes us an argument establishing that the doctrine is the best explanation of our intuitions. The appeal to intuitions is question-begging in this context.

## CONCLUSION

I believe that the data cited in this article give us a strong case for regarding the intuitions to which proponents of the doctrine of double effect appeal as too unreliable to play the role of justifying moral principles. It should be noted that there is some controversy surrounding the Knobe effect. Some philosophers and psychologists have offered rival interpretations of Knobe's data, some of which are less threatening to the doctrine of double effect than the view that he advances. McCann (2005), for instance, has argued that subjects distinguish between what is intended and what is done intentionally, while Guglielmo and Malle (2010) have suggested that Knobe's results are a product of the way in which Knobe framed his questions and do not reflect the deep structure of agents' judgments. I believe that these objections can be addressed without affecting the substance of the view I have advanced here. But addressing these objections would take us too far afield. Though it would be premature to regard the doctrine as decisively refuted, it should be clear that the data reviewed here presents a powerful challenge to a time-honored philosophical distinction.

In any case, the primary purpose of this paper is not to refute the doctrine of double effect but instead to illustrate the way in which neuroethics can help us to learn about, and thereby improve, the tools we use as ethicists. I also hoped to illustrate the interdisciplinary nature of neuroethics. Neuroethics is many things, and ought to welcome the work of philosophers and ethicists who wish to reflect on its subject matter, in just the same way as they might on bioethical dilemmas. But it also provides opportunities for researchers who aim to combine philosophical reflection with psychological and neuroscientific data, both to come to a better understanding of human agency and morality, and with a view to applying our new knowledge (for instance, in answering neuroethical questions). The reflections advanced here illustrate how we can utilize scientific data to hone our ethical tools.

Obviously, if these considerations show that the intuitions to which proponents of the doctrine of double effect appeal are unreliable, this is an important conclusion, since it would entail that we need to rethink some important moral issues (for instance, concerning the circumstances in which euthanasia is permissible). But my major purpose has not been to undermine the doctrine of double effect, but to illustrate how neuroethics, as an interdisciplinary endeavor, might proceed. Neuroethics so understood draws upon the expertise of psychologists, philosophers, neuroscientists, and other researchers, both to criticize existing moral principles and intuitions (as here), and to help to develop new moral principles or refine existing ones. Neuroethicists should continue to do applied ethics in the manner of bioethicists (for instance), but they should also work with investigators in the sciences of the mind, and utilize their results, in order to understand the tools they apply in assessing normative claims. We should use the neuroscience of ethics to illuminate the ethics of neuroscience. By doing so, we can produce better ethical theories and better justified normative conclusions, and contribute toward the great project of better understanding ourselves. ■

## REFERENCES

- Coady, C. A. J. 1992. *Testimony*. Oxford: Oxford University Press.
- Foot, P. 1978. The problem of abortion and the doctrine of the double effect. In *Virtues and vices*, 19–32. Oxford: Basil Blackwell.
- Greene, J., R. B. Sommerville, L. E. Nystrom, J. M. Darley, and J. D. Cohen. 2001. An fMRI investigation of emotional engagement in moral judgment. *Science* 293: 105–2108.
- Greene, J. 2003. From neural 'is' to moral 'ought': What are the moral implications of neuroscientific moral psychology? *Nature Reviews Neuroscience* 4: 847–850.
- Guglielmo, S., and B. F. Malle. 2010. Can unintended side-effects be intentional? Resolving a controversy over intentionality and morality. *Personality and Social Psychology Bulletin* 36: 1635–1647.
- Haidt, J. 2001. The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review* 108: 814–834.
- Haidt, J., S. H. Koller, and M. G. Dias. 1993. Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology* 65: 613–628.
- Hauser, M. 2006. *Moral minds: How nature designed our universal sense of right and wrong*. New York: HarperCollins.
- Kahane, G., and N. Shackel. 2008. Do abnormal responses show utilitarian bias? *Nature* 452(7185): E5.
- Knobe, J. 2003. Intentional action and side effects in ordinary language. *Analysis* 63: 190–193.
- Knobe, J. 2006. The concept of intentional action: A case study in the uses of folk psychology. *Philosophical Studies* 130: 203–231.
- Levy, N. 2006. The wisdom of the pack. *Philosophical Explorations* 9: 99–103.

- McCann, H. 2005. Intentional action and intending: Recent empirical studies. *Philosophical Psychology* 18: 737–748.
- Mikhail, J. Forthcoming. *Elements of moral cognition: Rawls' linguistic analogy and the cognitive science of moral and legal judgment*. Cambridge: Cambridge University Press.
- Roskies, A. 2002. Neuroethics for the new millenium. *Neuron* 35: 21–23.
- Schnall, S., J., Haidt, G. L., Clore, and A. H. Jordan. 2008. Disgust as embodied moral judgment. *Personality and Social Psychology Bulletin* 34: 1096–110
- Singer, P. 2005. Ethics and intuitions. *Journal of Ethics* 9: 331–352.
- Wheatley, T., and J. Haidt. 2005. Hypnotic disgust makes moral judgments more severe. *Psychological Science* 16: 780–784.