

Ethics // Spring 2026

Handout 3

Ethics and neuroscience: Greene

CONTRAST WITH HAIDT'S PROJECT. Greene begins with the major premiss shared by Haidt that there is a dual-process cognitive system, where one process is 'reasoning' and the other 'intuition' (see the summary in Haidt's Table 1). Haidt's conclusion was that our moral judgement is largely determined by intuitions, even if occasional exceptions are possible—e.g., with moral philosophers.

Crucially, however, Haidt did not distinguish between deontological and consequentialist judgements: his conclusion was meant to be applied across the board. Greene will argue that deontological judgements are emotional and 'intuitive', whereas consequentialist (including utilitarian) judgements are rational and 'cognitive' (where this term is technical). So the two cognitive tracks are occupied by two different moral judgements. Greene will come back to this disagreement with Haidt in page 63.

DEONTOLOGY AND CONSEQUENTIALISM. Greene has an interesting new suggestion. Normally, you'd think that deontology and consequentialism are two philosophical theories, in a sense *inventions*. So we might have:

(1) **Deontology:** the moral value of an action is determined by the agent's following certain unassailable principles.

Consequentialism: the moral value of an action is determined by its consequences with regard to a certain good like universal welfare (in utilitarianism).

But Greene rejects this natural thought. Instead, deontology and consequentialism are two patterns of responses to a morally relevant situation. They are intrinsic to our cognitive mechanisms. Since our concern is not metaphysics, let's downplay Greene's talk of essences and natural kinds. Let's simply say this: Although we may define and use the terms 'deontological' and 'consequentialist' arbitrarily as we please, in fact the persistent appeal of deontology and consequentialism is explained by the presence of those two cognitive patterns.

To sum up: Just like Haidt, we define deontology and consequentialism in terms of their 'characteristic judgements', i.e. verbal responses to a morally relevant situation. This division is not arbitrary: it reflects the underlying 'essences' of deontological and consequentialist philosophies (see Figure 1).

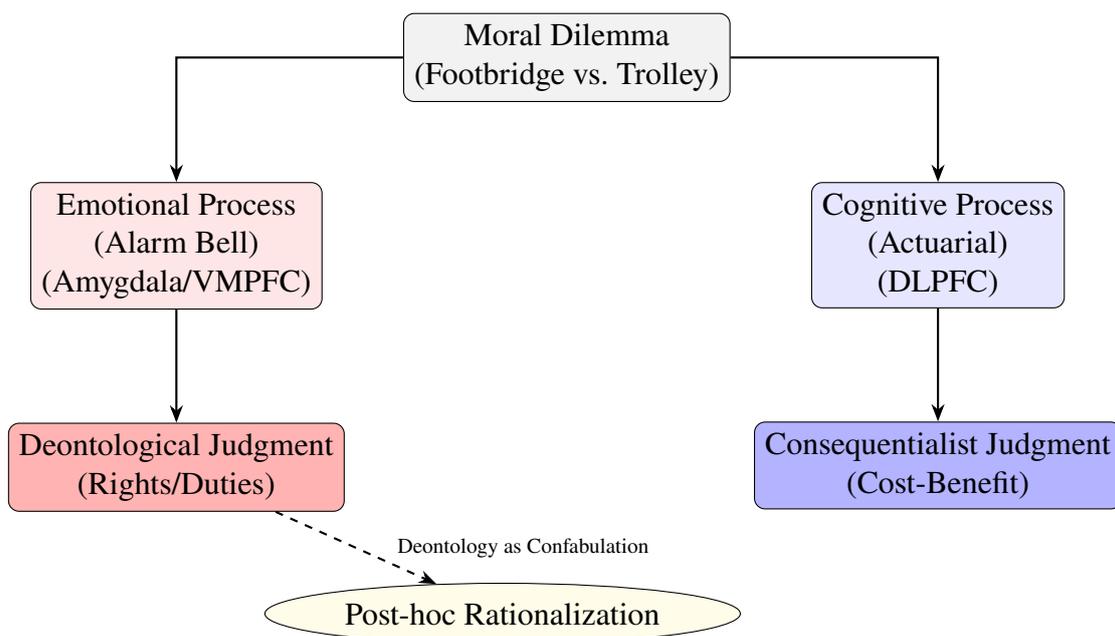


Figure 1: Main schema

COGNITION AND EMOTION. As Greene uses ‘cognition’, it is a behaviourally neutral representation triggering no particular automatic response on the part of the agent. ‘Emotion’, on the other hand, does trigger it. Further, they can be distinguished by their location in the brain: pre-frontal cortex for cognition, amygdala for emotion. 40

So there are four possibilities to consider: 41

- (a) Both deontological and consequentialist judgements are cognitive. (Rawls, Kohlberg)
- (b) Both deontological and consequentialist judgements are emotional. (Haidt)
- (c) Deontological judgements are cognitive, consequentialist judgements are emotional. (Hume?)
- (d) Deontological judgements are emotional, consequentialist judgements are cognitive. (Greene)

DIAGNOSING MORAL DILEMMAS. The puzzle to explain is the divergence we observe in the consensus over the Trolley dilemma and the Footbridge dilemma (see the setup in the text). Why do people think that it is OK, if not required, to divert the trolley even if this causes a man to die, but not OK to push the man with the same result? A popular philosophical explanation is that in the trolley case the man dies merely as a side effect. But in the footbridge case he is used directly as a means to avoid harming others—something that is morally inadmissible. 42

Yet we should consider the Loop case (details in the text) where the person is, again, used as a means, yet the consensus is that diverting the trolley then *is* permissible. So the philosophical explanation can’t be right.

Greene offers another explanation. The difference is the mode of violence. In the trolley and loop cases the violence exercised is remote and impersonal. In the footbridge case violence is salient. And when it is salient, an emotional response against it is triggered. That’s why we tend to judge this action unacceptable. In the footbridge case this response is not triggered. Thus there, the judgement is formed by ‘reasoning’, with the typical consequentialist reasons taking over. 43

This claim can be tested empirically. When you think of the trolley case, the areas of your brain associated with ‘cognition’ should light up. When you think of the footbridge case, the areas of your brain associated with ‘emotion’ should light up. That is what was observed in experiments. Similarly, it was shown that those who judged it permissible to push the man on the tracks took a longer time to respond than those who didn’t. This matches the idea that the former relied on reasoning, whilst the latter on emotion, with the consequentialist and deontology judgements elicited as a result. 44

REFLECTIONS. Is this a convincing argument? First, let’s tentatively quibble with Greene’s empirical results. When I require more time to come up with consequentialist responses, that on its own, might not mean much. Consequentialist calculations take longer, but the initial inclination to use such calculations (and therefore, to apply consequentialist reasoning at all) may well be quick.

This nuance aside, an alternative explanation of divergence may have to do with our intuitions about what constitutes an action. When I’m diverting the trolley, that’s what I am doing—diverting the trolley. This is an intuition about what *action* is being performed. No violence is involved in the action itself. When I push a man on the tracks, my intuition is that this is a ‘violent action’, not an ‘action *simpliciter*’. That’s how I classify it (in my brain?). So the alternative explanation is that we respond negatively to violent actions (where violence is an integral part of the action), but not to actions that merely *lead* to violence, as in the trolley case. I have intuitions about what constitutes *my* actions, and I distinguish them, still intuitively, from their consequences. The contrast, according to this explanation, is between violent/non-violent actions, rather than between salient violence/remote violence.

Another issue to consider is how familiar the context is. Consider a military commander sending a soldier to a mission extremely likely (virtually certain) to result in death. This is a case of using a person as mere means. Consider now the ‘Ayşegül Teyze case’ where we are asked to harvest organs from a useless person for the sake of a greater good. There again we use a person as a mere means. In the former case, the majority is OK with the action, not so in the latter case. Why? Well, it’s hard to diagnose this in terms of Greene’s emotional salience. In fact, the soldier case involves ‘up close and personal’ violence that the Ayşegül Teyze case does not. In addition to the action-definition explanation just given, there is also the familiarity of the soldier’s sacrifice, as opposed to the complete

strangeness of organ harvesting. The battle sacrifice has been with us forever, praised consistently and effusively. Organ harvesting is still in its infancy.

Similarly relevant are the precise descriptions of the actions. In one case we speak of ‘ sacrifice ’, ‘ courage ’ etc. These descriptions suggest praiseworthiness. In another, we speak of ‘ organ harvesting ’. If anything, this suggests something unsavoury and despicable.

BABY INTUITIONS. Greene claims support for his theory from other dilemmas as well. People were ambivalent about the *crying baby dilemma* (no uniform answers). But there was a consensus about the *infanticide dilemma* where they quickly judged that killing the baby is not permissible. Greene’s model predicts that that in both cases there is an emotional aversion to killing babies (i.e. aversion to up-close violence). Yet in the in the cb-dilemma there is also a strong cost-benefit, cognitive response going against the emotional response. Thus we have a ‘ response conflict ’. In some people the cognitive response dominates, in other the emotional response does (see Figure 2).

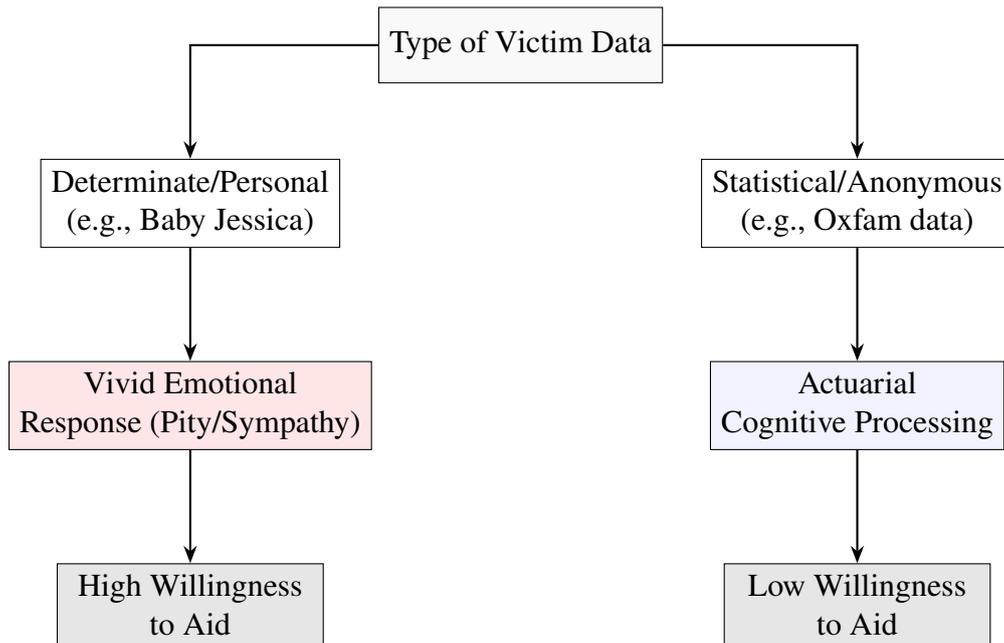


Figure 2: Victim differential

This claim can be empirically tested (generally, Greene claims the neuroscientific testability as a major virtue of his account). Indeed, the areas of the brain (the anterior cingulate cortex) associated with the response conflict have lit up when people were asked about the cb-dilemma. Also, these people’s brains showed increased activity in the areas associated with cognitive responses (in contrast to the people who were asked about infanticide).

SINGER’S CHALLENGE. In a hugely influential article, Singer argued that there is no difference between helping a child drowning up there in front of you and an anonymous child somewhere in India. We have a pair of dilemmas here:

(2) Proximate suffering: Are we obligated to help a drowning child (or adult!) here now, in front of you? The consensus, hopefully, is that we are.

Remote suffering: Are we obligated to help a starving child in India? At best there is no consensus, or likely there is a majority that we aren’t.

Singer’s challenge is that our responses in Remote suffering are untenable. The same grounds holding in Proximate suffering should hold in Remote suffering. Indeed, perhaps they are more convincing in the latter case, since saving a drowning child may involve a risk to yourself, but no such risk exists if you adopt *effective altruist* policies whereby you donate some of your income towards the welfare of Indian strangers. In short, so far as you find it obligatory to save the drowning child, you must also find obligatory to improve, in some sensible way, the welfare of remote strangers in India.

Remark 1. Singer’s consequentialism explicitly contradicts Mill’s classical utilitarianism.

Ignoring the validity of Singer’s consequentialism, Greene’s theory can explain the different attitudes we take toward the drowning child and Indian strangers. The former’s suffering is up close, and our emotions are kicking in. The latter’s problems are, by definition, anonymous, and emotions are idle. Only consistently rational utilitarians would adopt Singer’s position. Emotion-driven deontologists (so Greene) would permit ignoring the plight of Indian strangers, since there is no emotion to generate their principles. 47

But this example in particular seems to me problematic for Greene’s explanations. The drowning child elicits emotions in you, and your deontological principles become operative prohibiting you from carrying on indifferently. Suppose we accept this. But why can’t a deontologist still try to work out a principle to require our altruistic contributions to Indian strangers? We might imagine him saying:

- (3) If you help yourself and your family, but not the Indian, you are treating them not *equally*, not as equal humans.

Again, the deontologist might reframe the issue. He might claim that by refusing to donate to the Indian, you harm them. But harm is prohibited universally, hence you ought to donate. Therefore, Greene’s conclusion that our indifference to the Indian is determined in a ‘characteristically deontological way’ seems premature. 48

Greene has replies to this objection, however. Deontological judgement *may* become rational. But typically, it isn’t. 65

Further, even if it becomes rational, there is a suspect coincidence between quick emotional responses and deontological (ostensibly reasoned) judgement. Short of a Divinely pre-established harmony, there is no obvious explanation why our (empirically identifiable) emotional responses must track the rationally discoverable deontological moral truths. 69

APPROACHES TO PUNISHMENT. Greene’s lengthy discussion of punishment may be summarised rather quickly. The deontological idea of punishment is based on retribution, a kind of compensation for the wrong already done. The consequentialist idea is based on future effects, like promoting future obedience to law. Experiments show that people have predominantly deontological, emotionally charged ideas of punishment. Curiously, in one study people were instructed *explicitly* to think of punishment in a consequentialist way. But the subjects slapped punishments based on retribution, then added some more for deterrence (thus ostensibly complying with the consequentialist instructions)! See Figure 3. 50
51
53

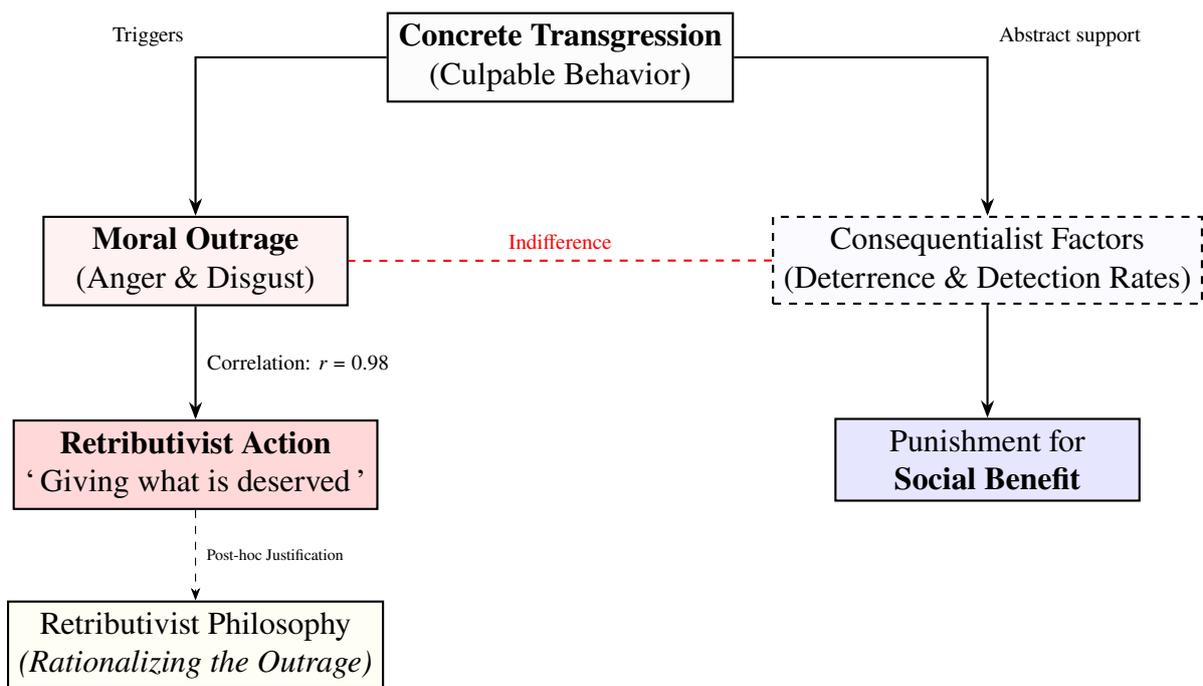


Figure 3: Punishment

Other than showing that deontological responses are driven by emotion (specifically, outrage), Greene’s model also predicts that the responses should vary when the subjects consider general 53

policies of punishment, and when they consider concrete situations and individuals. In the former case emotional responses are triggered less or not at all. In the latter case emotions (outrage) rule, since these are cases of ‘ up close violence ’. This is indeed confirmed by empirical studies.

HARMLESS ACTIONS. From the analysis of punishment it follows that the presence of harm is not a factor (or a decisive factor) in the deontological responses of the subjects. What of the absence of harm? Here, too, emotional responses to such actions were aligned with the weakened cognitive response. This is borne out further by looking at the differences among the populations. More Westernised populations were more reluctant to condemn these actions.

EVOLUTIONARY ROOTS OF ETHICS. Having surveyed the empirical grounds of deontology and consequentialism, Greene turns to broader questions. Why should deontological judgements and emotions go together? Well, why should there be moral emotions at all? Greene endorses an evolutionary premiss, that moral responses are evolutionary adaptations formed to deal with various situations of social life. Morality as a whole, as an institution, is an evolutionary adaptation. And we need emotions, rather than ‘ cognitions ’, because they are quick and reliable. See Figure 4.

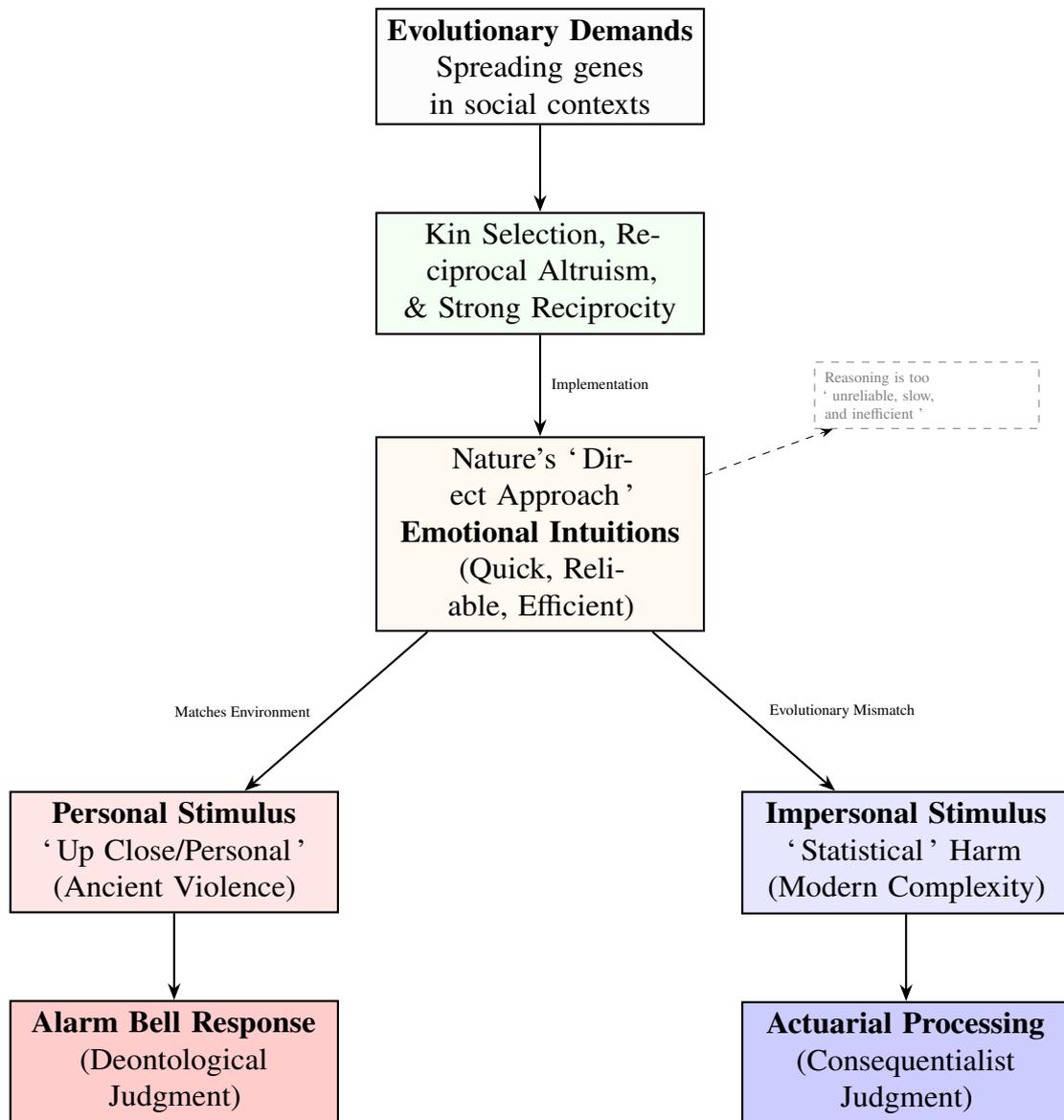


Figure 4: Evolutionary roles

Now, supposing that there are such moral emotions, why do they elicit specifically deontological philosophy? Greene embraces Haidt’s approach, that this philosophy is a rationalisation of the pre-loaded emotional responses. Such rationalisations, indeed, are observed in a wide variety of cases. Effectively, people are never content with the opaqueness and seeming arbitrariness of their ‘ quick ’

responses, and they always search for some (superficially) rational vindication of those responses. Compare the especially remarkable cases of confabulation.

62

In fact, that's how Greene proposes to see deontology: it is a moral confabulation. When presented with certain types of situation, like our own 'Ayşegül Teyze', the deontological thinker has a strong emotional response 'This can't be done!' However, just like the confabulators in other situations, they manufacture post-hoc rationalisations.

63

Remark 2. Observe a minor disagreement with Haidt. According to Haidt, moral confabulation is a function of the social character of morality (i.e. of moral judgement). Greene sees it as a result of a general tendency shared across different domains.

THE RATIONALITY OF CONSEQUENTIALISM. In a major break from Haidt and other similar-minded thinkers, Greene argues that consequentialism is immune to quick emotional responses. That's because, by its very nature, consequentialism is deliberative: we can't arrive at any consequentialist judgement before examining the details of the situation and weighing the effects of the action. Every such judgement is a 'complex guessing game'.

64



Greene acknowledges that consequentialism also involves emotions: weighing outcomes in terms of some quantity (like happiness) may well be driven by emotion. But they merely 'influence' this judgement, rather than 'dominate' it.

64

YSB