Lecture 07 : Moral Psychology

Stephen A. Butterfill
< s.butterfill@warwick.ac.uk >

Thursday, 29th February 2024

Contents

1. Linking Ethics to Moral Psychology: Dual-Process Theories

In the previous lecture, we were mostly concerned with the use of empirical claims about moral psychology *within* ethical arguments. Now we turn to whether discoveries about moral psychology can be used to undermine ethical arguments from the outside.

If you are short of time this week, go straight to *Greene contra Ethics (Railgun Remix)* (section §2), consider the outline of the argument and the implications. Then review sections associated with any of the premises you are unsure about. Optionally consider *Conclusion: Guesses Aren't Evidence* (section §7). Done.

1.1. Introduction

In this lecture, we will consider a loose reconstruction of Greene's argument for the claim (as I put it) that discoveries in moral psychology reveal that not-justified-inferentially premises about particular moral scenarios cannot be used in ethical arguments (Greene 2014).

If Greene is right, the methods of Foot (see *Foot and Trolley Cases: Kant Was Wrong* in Lecture 06), Kamm (see *Singer vs Kamm on Distance* in Lecture 06) and Thomson (see *Thomson's Other Method of Trolley Cases* in Lecture 06) are all misguided, along with many other philosophical arguments in ethics.

We will also eventually (but not in this lecture) examine Greene's further, logically independent contention that his argument supports the application of some kind of broadly consequentialist ethical theory to unfamiliar problems.

1.2. Reflective Equilibrium

If the argument introduced in this lecture is correct, it may support an objection to the method of reflective equilibrium when used in an attempt to discover ethical truths (Singer 2005).

This would be a significant result because reflective equilibrium 'is the dominant method in moral and political philosophy' (Knight 2023). Indeed, according to (Scanlon 2002, p. 149), reflective equilibrium is 'the best way of making up one's mind about moral matters' and 'it is the only defensible method: apparent alternatives to it are illusory.'

What is reflective equilibrium? Rawls introduces the idea like this:

> 'one may think of moral theory at first [...] as the attempt to describe our moral capacity [...] what is required is a formulation of a set of principles which, when conjoined to our beliefs and knowledge of the circumstances, would lead us to make these judgments with their supporting reasons were we to apply these principles conscientiously and intelligently' (Rawls 1999, p. 41; see Singer (1974) for critical discussion).

Roughly, then, the idea is to start with not-justified-inferentially judgements you are, on reflection, inclined to make.[1] And then to consider which principles might be consistent with these judgements. You may drop some of the judgements you start with depending on how well principles can be made to fit them.

1.3. Relation to Lecture 06

This lecture does not depend on *Lecture 06* (as I anticipate that you may skip one or the other of these lectures) but you may find it helpful to relate the two.

The key contrast is this: in *Lecture 06*, we were concerned with the use of empirical claims about moral psychology *within* ethical arguments. We considered attempts to show that moral psychology is relevant to ethics which rely on some philosophers' approaches being broadly correct. In this lecture, our concern is with whether discoveries in moral psychology can undermine the case for accepting non-empirical premises of ethical arguments *from the outside*. We will consider attempts to show that moral psychology is relevant to ethics which rely on some philosophers' approaches being substantially misguided.

To assist in understanding the contrast, a recap may be helpful ...

1.4. Recap of Lecture 06

Some arguments for ethical principles rely on noninferentially justified premises about particular moral scenarios. Among these arguments, some are straightforwardly undermined or supported by discoveries in moral psychology (see *Foot and Trolley Cases: Kant Was Wrong* in Lecture 06 and *Singer vs Kamm on Distance* in Lecture 06). Other arguments have no straightforward relation to discoveries in moral psychology (see *Thomson's Other*

[1] Compare Rawls (1951, p. 183): 'it is required that the judgement [...] not be determined by a systematic and conscious use of ethical principles.' Rawls goes on to motivate this requirement with the observation that 'We cannot test a principle honestly by means of judgments wherein it has been consciously and systematically used to determine the decision.'

Method of Trolley Cases in Lecture 06). Further, invoking discoveries about framing effects does not, by itself at least, appear to create significant challenges (see *Framing Effects: Emotion and Order of Presentation* in Lecture 06).

Greene's argument, if correct, shows that discoveries in moral psychology are, after all, relevant to evaluating Thomson's argument.

In Lecture 06, we considered framing effects (see *Framing Effects: Emotion and Order of Presentation* in Lecture 06). Greene's argument requires a deeper understanding of the processes underpinning ethical judgements than do arguments from framing effects.

2. Greene contra Ethics (Railgun Remix)

Do discoveries in moral psychology reveal that not-justified-inferentially premises about particular moral scenarios cannot be used in ethical arguments? This section outlines a loose reconstruction of one strand of Greene (2014)'s argument which, if successful, shows that the answer is yes.

Greene (2014)'s argument has been interpreted in a variety of ways, and has ambitious aims (including establishing that a broadly consequentialist theory is preferable to any deontological theory). Since Greene's argument has been the target of several objections, our strategy will be first to consider whether we can craft a loose reconstruction of one strand of the argument which aims to establish a conclusion more modest than Greene's own (although one with interesting implications). If that succeeds, we may then consider whether further arguments for Greene's more ambitious conclusions succeed.

2.1. Argument Outline

1. Ethical judgements are explained by a dual-process theory, which distinguishes faster from slower processes (see *A Dual Process Theory of Ethical Judgement* (section §3)).
2. Faster processes are unreliable in unfamiliar situations (see *Cognitive Miracles: When Are Fast Processes Unreliable?* (section §4)).
3. Therefore, we should not rely on faster process in unfamiliar situations [from 2].
4. When philosophers rely on not-justified-inferentially premises, they are relying on faster processes (see *What Is the Role of Fast Processes In Not-Justified-Inferentially Judgements?* (section §5)).

5. We have reason to suspect that the moral scenarios philosophers consider are unfamiliar situations (see *Which Moral Scenarios Are Unfamiliar?* (section §6)).

6. Therefore, not-justified-inferentially premises about particular moral scenarios cannot be used in ethical arguments where the aim is to establish knowledge of their conclusions [from 3, 4 and 5].

2.2. Implications

The above argument implies that Thomson's method of trolley cases is misguided (see *Thomson's Other Method of Trolley Cases* in Lecture 06), along with many other philosophical arguments in ethics.

The above argument, if successful, also implies the falsity of Audi's view about ethics:

> 'Episodic intuitions [...] can serve as data [...] ... beliefs that derive from them receive prima facie justification' (Audi 2015, p. 65).

The above argument does not favour one type (e.g. deontological vs consequentialist) of ethical theory, nor one approach to doing ethics (e.g. case-based vs systematic).[2] (We will eventually consider whether further arguments succeed in establishing either such favouritism.)

The above argument does not imply that philosophers should give up on arguments involving not-justified-inferentially premises about particular moral scenarios. Aristotelian theories of the physical, although much less useful than the successors which arose when scientists moved away from reliance on not-justified-inferentially premises, remain useful in some situations. And in the cases of ethics, there may be no better alternative approach.

The above argument implies that when using arguments involving not-justified-inferentially premises about particular moral scenarios (as in *Thomson's Other Method of Trolley Cases* in Lecture 06, for example), the aim should not be to establish knowledge of their conclusions. Instead it might be to characterise aspects of moral cognition (as Kozhevnikov & Hegarty (2001) use an Aristotelian theory of the physical to characterise physical cognition).

[2] The loose reconstruction may appear to favour systematic over case-based approaches to ethics because its conclusion concerns judgements about particular moral scenarios. This appearance is misleading. The conclusion is framed in this way for simplicity. The argument can be straightforwardly generalised to cover not-justified-inferentially premises about moral principles too.

Or the aim might be to understand what consistency with certain judgements would require.

2.3. Alternative Reconstructions

Kumar & Campbell (2012) provide an alternative reconstruction of Green's argument (which, helpfully, is a refinement on a critique of Berker (2009)'s earlier reconstruction: Kumar and Campbell are probably easier to understand). They analyse Greene's argument as a debunking argument. This means that (a) it depends on premises about which factors are morally relevant; and (b) is is open to the response that facts about which factors explain judgements are ethically irrelevant (see Rini 2017, 1443[3]).

Why bother with my loose reconstruction when I could just borrow Kumar & Campbell (2012)'s? While their reconstruction may be more faithful to the original (Greene 2014), my loose reconstruction does not depend on premises about which factors are morally relevant nor does it require the premises that facts about which factors explain why certain judgements are made are ethically relevant. This enables the loose reconstruction to avoid some objections (see *Quick Objections to Greene's Argument* in Lecture 08).

3. A Dual Process Theory of Ethical Judgement

Dual process theories of moral psychology claim that moral abilities involve two, or more, processes that are independent and divergent. They are independent in the sense that the conditions which influence whether they occur and which outputs they generate do not completely overlap. And they are divergent in the sense their outputs can conflict (given a single scenario, one process' output may imply rightness whereas the another's implies wrongness).

3.1. Please Answer This Dilemma First

The recording and slides make use of the following dilemma. Please answer it before you start.

[3] In this passage, Rini cites Nagel (1997, p. 105) in support of the view that discoveries about moral psychology cannot 'change our moral beliefs'. Note that the paragraph she cites from ends with a much weaker claim opposing 'any blanket attempt to displace, defuse, or subjectivize' moral concerns. Further, Nagel's essay starts with the observation that moral reasoning 'is easily subject to distortion by morally irrelevant factors ... as well as outright error' (Nagel 1997, p. 101). So while one of Nagel's assertions supports Rini's interpretation, it is unclear to me that Rini is right about Nagel's considered position. But I could easily be wrong.

> 'You are part of a group of ecologists who live in a remote stretch of jungle. The entire group, which includes eight children, has been taken hostage by a group of paramilitary terrorists. One of the terrorists takes a liking to you. He informs you that his leader intends to kill you and the rest of the hostages the following morning.
>
> 'He is willing to help you and the children escape, but as an act of good faith he wants you to kill one of your fellow hostages whom he does not like. If you refuse his offer all the hostages including the children and yourself will die. If you accept his offer then the others will die in the morning but you and the eight children will escape.
>
> 'Would you kill one of your fellow hostages in order to escape from the terrorists and save the lives of the eight children?' (Koenigs et al. 2007)

3.2. The Stripped-Down Dual-Process Theory

According to this theory:

> Two (or more) ethical processes are distinct in this sense: the conditions which influence whether they occur, and which outputs they generate, do not completely overlap.
>
> One process is faster than another: it makes fewer demands on scarce cognitive resources such as attention, inhibitory control and working memory.

A key feature of the stripped-down dual-process theory is its *theoretical modesty*: it involves minimal commitments concerning the particular characteristics of the processes. Identifying characteristics of the process is a matter of discovery.

3.3. Why This Dual-Process Theory?

Greene offers an elaborate dual-process theory of ethical cognition, one which incorporates controversial claims about consequentialism and emotion.[4] As these claims are neither essential features of a dual-process theory nor necessary for the overall argument we are developing (see *Greene contra*

[4] See Paxton & Greene 2010 for a compact overview of Greene's theory. The theory has been presented in a variety of different ways (see, for example, Cushman et al. (2010) for an alternative presentation).

Ethics (Railgun Remix) (section §2)), we may consider a stripped-down dual process theory instead.

3.4. What Does the Dual-Process Theory Predict?

To make use of existing evidence, we have to add an auxiliary assumption to the dual-process theory:

> The slow process is responsible for characteristically consequentialist responses; the fast for other responses.[5]

Prediction 1: Increasing cognitive load will selectively slow consequentialist responses. This prediction has been confirmed (Greene et al. 2008).

Prediction 2: Limiting the time available to make a decision will reduce consequentialist responses. This prediction also appears to have been confirmed:

> 'The model detected a significant effect of time pressure, p = .03 (see Table 1), suggesting that the slope of utilitarian responses was steeper for participants under time pressure. [...] participants under time pressure gave less utilitarian responses than control participants to scenarios featuring low kill–save ratios, but reached the same rates of utilitarian responses for the highest kill–save ratios' (Trémolière & Bonnefon 2014, p. 927).[6]

On the face of it, then, the dual-process theory appears well supported by evidence (and Greene 2014 cites much further evidence). We may therefore accept it for now.

Of course you will need to evaluate the evidence properly (for general guidance on evaluating evidence, see *Moral Intuitions and Emotions: Evaluating the Evidence* in Lecture 02) before you can claim to know whether or not the dual-process theory is true. We will consider some more evidence for, and against, the dual-process theory next week (look out especially for process dissociation).

3.5. Other Dual-Process Theories of Ethical Cognition

Dual-process theories of ethical cognition are widely endorsed but come in many varieties. All of the following are elaborations of the stripped-down dual-process theory above.

---

[5] Greene (2014) suggests that the fast process is responsible for characteristically deontological responses, but this further assumption is not required to generate the predictions considered here.

[6] Later we will consider an alternative interpretation of the same findings due to Gawronski et al. (2018, p. 1006).

3.5.1. Cushman's Dual-Process Theory

Cushman supplements the core idea of a dual-process theory with a distinction between model-free and model-based learning:

> 'the functional role of value representation in a model-free system is to select actions without any knowledge of their actual consequences, whereas the functional role of value representation in a model-based system is to select actions precisely in virtue of their expected consequences' (Cushman 2013, p. 285).

Cushman also proposes that a good dual-process theory should explain patterns of judgements on dilemmas like the trolley problems, and that this requires appeal to distinction between model-free and model-based:

> 'It is the contrast between model-free and model-based systems—or between action- and outcome-based valuation—that can explain the conflict engendered by moral dilemmas' (Cushman 2013, p. 285).

We consider Cushman's proposal in *Appendix: Dual Process Theory and Auxiliary Hypotheses* in Lecture 08, where a quarter of it is adopted.

3.5.2. Kumar's Dual-Process Theory

Kumar proposes what he calls a 'minimalist' dual-process theory:

> 'A minimalist model says that two types of processes generate moral judgements. Type 1 processes are fast, spontaneous, unconscious, and involve emotional processing; type 2 processes are slow, controlled, conscious, and involve reasoned processing. In short, some moral judgements arise as a flash of feeling, while others issue from conscious deliberation' (Kumar 2016, p. 791).

This is not very different from the stripped-down theory I offered. So why not just use Kumar's theory rather than making my own?[7] In my view, it is not minimalist enough: there is little evidence on consciousness, control or speed; nor do we know enough about the phenomenology to postulate 'a flash of feeling'.

Interestingly, Kumar does not accept Greene's proposed link between fast processes and characteristically consequentialist responses. He does, however, link the fast processes to emotions:

[7] The stripped-down dual-process theory not strictly speaking mine. It was developed jointly with Ian Apperly, Jason Low and Hannes Rakoczy.

> 'Each process type gives rise to one element of moral judgement: type 1 processes generate moral emotions and type 2 processes generate moral beliefs. A minimalist model also explains conflict cases' (Kumar 2016, p. 792).

Defending this claim requires finding some evidence for the link. My view is that while emotion probably plays different roles in fast and slow processes, it is likely to feature in both (as Cushman 2013 suggests). For this reason I take Kumar's bet on fast/slow linking to emotions/beliefs to be risky. And it is not required to generate predictions currently being tested.

3.5.3. Haidt's Dual-Process Theory

Haidt's Social Intuitionist Model of Moral Judgement (which is a part of Moral Foundations Theory; see *Moral Foundations Theory: An Approach to Cultural Variation* in Lecture 04) could be interpreted as a kind of dual-process theory because it distinguishes intuition and reasoning as two kinds of process.

As we saw, the Social Intuitionist Model of Moral Judgement involves a variety of further claims—such as that 'moral reasoning is done primarily for socially strategic purposes' (Graham et al. 2013, p. 66)—which are not essential features of a dual-process theory.[8]

3.6. Two Systems?

Although two systems theories are sometimes understood as making claims over and above those of a dual-process theory (e.g. Gawronski et al. 2014), others do not make any distinction:

> 'We use the term "system" only as a label for collections of cognitive processes that can be distinguished by their speed, their controllability, and the contents on which they operate' (Kahneman & Frederick 2005, p. 267).

[8] Paxton & Greene (2010, pp. 513–4) offer a concise comparison: 'there are two critical differences between Haidt's SIM and Greene's dual-process model. First, while the SIM posits that reasoned judgment within an individual is, ' 'rare, occurring primarily in cases in which the intuition is weak and processing capacity is high," Greene's dual-process model allows that moral reasoning—especially utilitarian/consequentialist reasoning—may be a ubiquitous feature of moral common sense. Second, according to the SIM, social influence on moral judgment only occurs when one person succeeds in modifying another's intuition.'

4. Cognitive Miracles: When Are Fast Processes Unreliable?

Fast processes are unreliable when deployed to solve unfamiliar problems. But if (as I suppose) we do not know much about how the fast processes work in the case of ethics, we cannot know which problems are unfamiliar. Can we nevertheless make practical use of the principle that fast processes are unreliable when deployed to solve unfamiliar problems?

4.1. Cognitive Miracles

In what situations could fast processes yield correct responses?

> 'genetic transmission, cultural transmission, and learning from personal experience [...] are the only mechanisms known to endow [fast] processes with the information they need to function well' (Greene 2014, p. 715).

> 'it would be a cognitive miracle if we had reliably good moral instincts about unfamiliar moral problems' (Greene 2014, p. 715).

> 'The *No Cognitive Miracles Principle*: When we are dealing with unfamiliar* moral problems, we ought to rely less on [...] automatic emotional responses and more on [...] conscious, controlled reasoning, lest we bank on cognitive miracles' (Greene 2014, p. 715).

4.2. How to Apply the No Cognitive Miracles Principle?

It is tricky to apply this principle. For instance, is how to win a chess match an unfamiliar problem?

Although it may initially seem reasonable to speculate that how to win a chess match is an unfamiliar problem, expert chess players are supposed to rely on faster processes.

Are cartoons unfamiliar situations to someone who has never seen one? Although it may initially seem reasonable to speculate that they are (humans presumably encountered few 2D schematic animations in evolution), fast processes appear to have no problem with them. Why? Because the fast processes that underpin physical cognition are driven by principles, and these principles (although false) can be applied to new situations.

Because of the way we defined an unfamiliar problem, knowing whether a problem is unfamiliar typically depends on knowing something about the

structure of the fast processes. Which, arguably, we do not in the case of ethics.

Does this mean the No Cognitive Miracles Principle is useless? Not at all. There are at least two ways we might apply it in practice even without knowing which situations are unfamiliar.

4.3. Wicked Learning Environments

Hogarth (among many researchers) has studied when fast processes can be reliably used even in the absence of knowing in detail how they work. (This is a practical problem in many areas of life.) He concludes:

> 'When a person's past experience is both representative of the situation relevant to the decision and supported by much valid feedback, trust the intuition; when it is not, be careful' (Hogarth 2010, p. 343; see Kahneman & Klein 2009, p. 520 for a related view).

This suggests a practical way to avoid relying on cognitive miracles even without knowing exactly which situations and problems are unfamiliar.

But this is not the only way to avoid relying on cognitive miracles.

4.4. Disagreements

Greene argues that it is reasonable to suppose that where there is fully informed disagreement about what to do, we are likely to be in an unfamiliar situation:

> 'we can use disagreement as a proxy for lack of familiarity*. If two parties have a practical moral disagreement—a disagreement about what to do, not about why to do it—it's probably because they have conflicting intuitions. This means that, from a moral perspective, if not from a biological perspective, at least one party's automatic settings are going astray. (Assuming that both parties have adequate access to the relevant nonmoral facts.) Absent a reliable method for determining whose automatic settings are misfiring, both parties should distrust their intuitions' (Greene 2014, p. 716).

Greene (2017) provides further discussion relevant to the question of which situations are, or might reasonably be suspected of being, unfamiliar.

4.5. Which comparison: Linguistic or Physical?

The slides and recording use a comparison between ethical and physical cognition. This assists in arriving at the view that fast processes are sometimes unreliable.

How would things look if instead we compared ethical to linguistic cognition? Roughly speaking, the facts in linguistics are determined by how the fast processes operate together with some collective cultural decision-making. It is hard even to make sense of the idea that the fast processes are unreliable because the linguistic facts are overwhelmingly determined by how the fast processes operate. (The collective cultural decision-making is a relatively recent, and relatively superficial, phenomenon.)

To illustrate, Jackendoff (2003, p. 19) observes that a postulation about the syntactic structure of a sentence 'is to be treated as a model of something in the mind of a speaker of English who says or hears this sentence.' In other words, the target is a fact about how the fast process operates. Claims of unreliability are therefore limited to cases where competence and performance come apart; it is not possible that a linguistic theory could discover that fast processes embody a systematically distorted view of the linguistic.

Is the comparison with the linguistic plausible? Relying on it would likely commit you to quite a dismal view of ethics (see *Lecture 04*, particularly *Moral Pluralism: Beyond Harm* in Lecture 04).

5. What Is the Role of Fast Processes In Not-Justified-Inferentially Judgements?

Philosophy is thinking in slow motion (Campbell). But then how could fast processes be relevant? Fast processes have little or no *direct* influence over non-justified-inferentially judgements. Despite this, they may dominate through *indirect* influence where knowledge is absent. For fast processes give rise to appearances (and high subjective confidence). These appearances provide material for reflection. In the absence of knowledge, reflection on how things appear is likely to determine how you judge them to be. In this way, fast processes can dominate, albeit indirectly, even glacial not-justified-inferentially judgements.

6. Which Moral Scenarios Are Unfamiliar?

There are at least two reasons to suspect that the moral scenarios philosophers typically consider are unfamiliar situations.

Do we have reason to suspect that the moral scenarios philosophers typically consider are unfamiliar situations?

6.1. Reason 1: Philosophical Methods

Even on the view most charitable to the argument's likely opponents (e.g. Railton 2014), some moral scenarios will be bizarre enough to count as unfamiliar. Although we do not know which these are (as far as I can tell), philosophers' interest in fine distinctions and edge cases increases the probability of hitting on unfamiliar situations.[9]

6.2. Reason 2: Signature Limits

We also know that there fast processes in other domains exhibit a range of signature limits even in adults and are unaffected by expertise, including:

- object cognition (Kozhevnikov & Hegarty 2001)
- mindreading (Low et al. 2016)
- number cognition (Feigenson et al. 2004)

This is no accident. Any broadly inferential process must make a trade-off between speed and accuracy. As more than a century of cognitive science has found (Henmon 1911; Link & Tindall 1971; Heitz 2014).[10]

Consequently, even for experts with much experience, some quite ordinary-seeming scenarios may be turn out to be unfamiliar. We should therefore be suspicious that at least some moral scenarios philosophers consider will turn out to involve signature limits, which would make them unfamiliar.

6.3. Going Deeper

Greene (2017) takes up the topic in detail.

[9] This may be a virtue of philosophical practice. Comparison with the physical case indicates that considering what turn out to be unfamiliar situations may be important for making discoveries (at least, Moletti (2000, p. 147) seems justifiably excited about vertical motion).

[10] To illustrate, suppose you were required to judge which of two only very slightly different lines was longer. All other things being equal, making a faster judgement would involve being less accurate, and being more accurate would require making a slower judgement. (This idea is due to Henmon (1911), who has been influential although he didn't actually get to manipulate speed experimentally because of 'a change of work' (p.~195).)

6.4. Which comparison: Linguistic or Physical?

The slides and recording use a comparison between ethical and physical cognition. This assists in arriving at the two reasons above.

How would things look if instead we compared ethical to linguistic cognition? As we saw in *Cognitive Miracles: When Are Fast Processes Unreliable?* (section §4), on any standard view it is not possible that a linguistic theory could discover that fast processes embody a systematically distorted view of the linguistic. One consequence is that is no easy way to make sense of the idea that there could be unfamiliar problems in the linguistic domain. So accepting the comparison with linguistic cognition might well lead us to reject this premise of the argument and deny that we would have any reason to suspect that the moral scenarios philosophers typically consider are unfamiliar situations.

Would accepting the comparison with linguistic cognition allow us to defend some proposed methods for gaining ethical knowledge such as Foot's, Kamm's or *Thomson's Other Method of Trolley Cases* in Lecture 06?

While accepting the comparison with linguistic cognition would mean that philosophers can avoid the conclusion of the loose reconstruction of Greene (2014)'s argument, it leads to a distinct, no less pressing challenge.

In linguistics, there is growing awareness that it is a mistake to rely on expert judgements (see, for example Wasow & Arnold 2005; Gibson & Fedorenko 2010; and Dąbrowska 2010). Understanding how fast linguistic processes work requires careful experiment, not introspective guesswork. Similar considerations apply in the case of ethics.

Therefore, even if we accept the comparison with linguistic cognition, we can still reach a conclusion that is close to, and has much the same implications for ethics as, the conclusion of the loose reconstruction of Greene (2014)'s argument:

> [alternative conclusion] Premises about judgements about particular moral scenarios need to be supported by carefully controlled experiments if they are to be used in ethical arguments where the aim is to establish knowledge of their conclusions.

7. Conclusion: Guesses Aren't Evidence

Discoveries in moral psychology reveal that not-justified-inferentially premises about particular moral scenarios cannot be used in ethical arguments insofar as the arguments aim to establish knowledge of their conclusions.

We have been exploring whether a loose reconstruction of an argument (as outlined in *Greene contra Ethics (Railgun Remix)* (section §2)) succeeds in establishing that not-justified-inferentially premises about particular moral scenarios cannot be used in ethical arguments insofar as the arguments aim to establish knowledge of their conclusions.

It does.

Glossary

automatic As we use the term, a process is *automatic* just if whether or not it occurs is to a significant extent independent of your current task, motivations and intentions. To say that *mindreading is automatic* is to say that it involves only automatic processes. The term 'automatic' has been used in a variety of ways by other authors: see Moors (2014, p. 22) for a one-page overview, Moors & De Houwer (2006) for a detailed theoretical review, or Bargh (1992) for a classic and very readable introduction 18

characteristically consequentialist According to Greene, a judgement is *characteristically consequentialist* (or *characteristically utilitarian*) if it is one in 'favor of characteristically consequentialist conclusions (eg, "Better to save more lives")' (Greene 2007, p. 39). According to Gawronski et al. (2017, p. 365), 'a given judgment cannot be categorized as [consequentialist] without confirming its property of being sensitive to consequences.' 9, 10

characteristically deontological According to Greene, a judgement is *characteristically deontological* if it is one in 'favor of characteristically deontological conclusions (eg, "It's wrong despite the benefits")' (Greene 2007, p. 39). According to Gawronski et al. (2017, p. 365), 'a given judgment cannot be categorized as deontological without confirming its property of being sensitive to moral norms.' 9

cognitively efficient A process is *cognitively efficient* to the degree that it does not consume working memory and other scarce cognitive resources. 18

David 'David is a great transplant surgeon. Five of his patients need new parts—one needs a heart, the others need, respectively, liver, stomach, spleen, and spinal cord—but all are of the same, relatively rare, blood-type. By chance, David learns of a healthy specimen with that very blood-type. David can take the healthy specimen's parts, killing him,

and install them in his patients, saving them. Or he can refrain from taking the healthy specimen's parts, letting his patients die' (Thomson 1976, p. 206). 20

debunking argument A *debunking argument* aims to use facts about why people make a certain judgement together with facts about which factors are morally relevant in order to undermine the case for accepting it. Königs (2020, p. 2607) provides a useful outline of the logic of these arguments (which he calls 'arguments from moral irrelevance'): 'when we have different intuitions about similar moral cases, we take this to indicate that there is a moral difference between these cases. This is because we take our intuitions to have responded to a morally relevant difference. But if it turns out that our case-specific intuitions are responding to a factor that lacks moral significance, we no longer have reason to trust our case-specific intuitions suggesting that there really is a moral difference. This is the basic logic behind arguments from moral irrelevance' (Königs 2020, p. 2607). 7

dual-process theory Any theory concerning abilities in a particular domain on which those abilities involve two or more processes which are distinct in this sense: the conditions which influence whether one mindreading process occurs differ from the conditions which influence whether another occurs. 8, 10, 11

Edward 'Edward is the driver of a trolley, whose brakes have just failed. On the track ahead of him are five people; the banks are so steep that they will not be able to get off the track in time. The track has a spur leading off to the right, and Edward can turn the trolley onto it. Unfortunately there is one person on the right-hand track. Edward can turn the trolley, killing the one; or he can refrain from turning the trolley, killing the five' (Thomson 1976, p. 206). 20

fast A *fast* process is one that is to to some interesting degree cognitively efficient (and therefore likely also some interesting degree automatic). These processes are also sometimes characterised as able to yield rapid responses.

Since automaticity and cognitive efficiency are matters of degree, it is only strictly correct to identify some processes as faster than others.

The fast-slow distinction has been variously characterised in ways that do not entirely overlap (even individual author have offered differing characterisations at different times; e.g. Kahneman 2013; Morewedge & Kahneman 2010; Kahneman & Klein 2009; Kahneman 2002): as its

advocates stress, it is a rough-and-ready tool rather than an element in a rigorous theory. 5, 8, 11–15, 19

loose reconstruction (of an argument). A reconstruction which prioritises finding a correct argument for a significant conclusion over faithfully representing the argument being reconstructed. 3, 5, 7

Moral Foundations Theory The theory that moral pluralism is true; moral foundations are innate but also subject to cultural learning, and the Social Intuitionist Model of Moral Judgement is correct (Graham et al. 2019). Proponents often claim, further, that cultural variation in how these innate foundations are woven into ethical abilities can be measured using the Moral Foundations Questionnare (Graham et al. 2009; Graham et al. 2011). Some empirical objections have been offered (Davis et al. 2016; Davis et al. 2017; Doğruyol et al. 2019). See **??**. 11

not-justified-inferentially A claim (or premise, or principle) is not-justified-inferentially if it is not justified in virtue of being inferred from some other claim (or premise, or principle).

Claims made on the basis of perception (*That jumper is red*, say) are typically not-justified-inferentially.

Why not just say 'noninferentially justified'? Because that can be read as implying that the claim *is* justified, noninferentially. Whereas 'not-justified-inferentially' does not imply this. Any claim which is not justified at all is thereby not-justified-inferentially. 3–6, 16, 17

reflective equilibrium A method that is supposed to provide justification for claims. The idea is to gather considered judgements about particular situations and attempt to identify principles which from which those judgements could be inferred, and then to adjust the judgements and principles so that they cohere. The canonical statement is Rawls (1999) (but Rawls 1951 is a useful earlier statement). Authoritative secondary sources are Knight (2023) and Scanlon (2002). 3

signature limit A *signature limit* of a system is a pattern of behaviour the system exhibits which is both defective given what the system is for and peculiar to that system. A *signature limit* of a model is a set of predictions derivable from the model which are incorrect, and which are not predictions of other models under consideration. 15

slow converse of fast. 5, 11

Social Intuitionist Model of Moral Judgement A model on which intuitive processes are directly responsible for moral judgements (Haidt & Bjorklund 2008). One's own reasoning does not typically affect one's own moral judgements, but (outside philosophy, perhaps) is typically used only to provide post-hoc justification after moral judgements are made. Reasoning does affect others' moral intuitions, and so provides a mechanism for cultural learning. 11, 19

trolley problem 'Why is it that Edward may turn that trolley to save his five, but David may not cut up his healthy specimen to save his five?' (Thomson 1976, p. 206). 10

unfamiliar problem An unfamiliar problem (or situation) is one 'with which we have inadequate evolutionary, cultural, or personal experience' (Greene 2014, p. 714). 3, 5, 6, 12–16

References

Audi, R. (2015). Intuition and Its Place in Ethics. *Journal of the American Philosophical Association*, *1*(1), 57–77.

Bargh, J. A. (1992). The Ecology of Automaticity: Toward Establishing the Conditions Needed to Produce Automatic Processing Effects. *The American Journal of Psychology*, *105*(2), 181–199.

Berker, S. (2009). The Normative Insignificance of Neuroscience. *Philosophy & Public Affairs*, *37*(4), 293–329.

Cushman, F. (2013). Action, Outcome, and Value: A Dual-System Framework for Morality. *Personality and Social Psychology Review*, *17*(3), 273–292.

Cushman, F., Young, L., & Greene, J. D. (2010). Multi-system moral psychology. In J. M. Doris, M. P. R. Group, et al. (Eds.), *The moral psychology handbook* (pp. 47–71). Oxford: OUP.

Dąbrowska, E. (2010). Naive v. expert intuitions: An empirical study of acceptability judgments. *The Linguistic Review*, *27*(1), 1–23.

Davis, D., Dooley, M., Hook, J., Choe, E., & McElroy, S. (2017). The Purity/Sanctity Subscale of the Moral Foundations Questionnaire Does Not Work Similarly for Religious Versus Non-Religious Individuals. *Psychology of Religion and Spirituality*, *9*(1), 124–130.

Davis, D., Rice, K., Tongeren, D. V., Hook, J., DeBlaere, C., Worthington, E., & Choe, E. (2016). The Moral Foundations Hypothesis Does Not Replicate Well in Black Samples. *Journal of Personality and Social Psychology*, *110*(4).

Doğruyol, B., Alper, S., & Yilmaz, O. (2019). The five-factor model of the moral foundations theory is stable across WEIRD and non-WEIRD cultures. *Personality and Individual Differences*, *151*, 109547.

Feigenson, L., Dehaene, S., & Spelke, E. S. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*(7), 307–314.

Gawronski, B., Armstrong, J., Conway, P., Friesdorf, R., & Hütter, M. (2017). Consequences, norms, and generalized inaction in moral dilemmas: The CNI model of moral decision-making. *Journal of personality and social psychology*, *113*(3), 343–376.

Gawronski, B., Conway, P., Armstrong, J., Friesdorf, R., & Hütter, M. (2018). Effects of incidental emotions on moral dilemma judgments: An analysis using the CNI model. *Emotion*, *18*(7), 989–1008.

Gawronski, B., Sherman, J. W., & Trope, Y. (2014). Two of what? a conceptual analysis of dual-process theories. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual-process theories of the social mind* (pp. 3–19). New York: Guilford Press.

Gibson, E. & Fedorenko, E. (2010). Weak quantitative standards in linguistics research. *Trends in Cognitive Sciences*, *14*(6), 233–234.

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral Foundations Theory: The Pragmatic Validity of Moral Pluralism. In P. Devine & A. Plant (Eds.), *Advances in Experimental Social Psychology*, volume 47 (pp. 55–130). Academic Press.

Graham, J., Haidt, J., Motyl, M., Meindl, P., Iskiwitch, C., & Mooijman, M. (2019). Moral Foundations Theory: On the advantages of moral pluralism over moral monism. In K. Gray & J. Graham (Eds.), *Atlas of Moral Psychology*. New York: Guilford Publications.

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, *96*(5), 1029–1046.

Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology*, *101*(2), 366–385.

Greene, J. D. (2007). The Secret Joke of Kant's Soul. In W. Sinnott-Armstrong (Ed.), *Moral Psychology, Vol. 3* (pp. 35–79). MIT Press.

Greene, J. D. (2014). Beyond Point-and-Shoot Morality: Why Cognitive (Neuro)Science Matters for Ethics. *Ethics*, *124*(4), 695–726.

Greene, J. D. (2017). The rat-a-gorical imperative: Moral intuition and the limits of affective learning. *Cognition*, *167*, 66–77.

Greene, J. D., Morelli, S. A., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition*, *107*(3), 1144–1154.

Haidt, J. & Bjorklund, F. (2008). Social intuitionists answer six questions about moral psychology. In W. Sinnott-Armstrong (Ed.), *Moral Psychology, Vol 2: The Cognitive Science of Morality: Intuition and Diversity* chapter 4, (pp. 181–217). Cambridge, Mass: MIT press.

Heitz, R. P. (2014). The speed-accuracy tradeoff: history, physiology, methodology, and behavior. *Decision Neuroscience*, *8*, 150.

Henmon, V. A. C. (1911). The relation of the time of a judgment to its accuracy. *Psychological Review*, *18*(3), 186–201.

Hogarth, R. M. (2010). Intuition: A Challenge for Psychological Research on Decision Making. *Psychological Inquiry*, *21*(4), 338–353.

Jackendoff, R. (2003). *Foundations of Language: Brain, Meaning, Grammar, Evolution.* Oxford, New York: Oxford University Press.

Kahneman, D. (2002). Maps of bounded rationality: A perspective on intuitive judgment and choice. In T. Frangsmyr (Ed.), *Le Prix Nobel, ed. T. Frangsmyr, 416–499.*, volume 8 (pp. 351–401). Stockholm, Sweden: Nobel Foundation.

Kahneman, D. (2013). *Thinking, Fast and Slow.* New York: Farrar, Straus and Giroux.

Kahneman, D. & Frederick, S. (2005). A model of heuristic judgment. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 267–293). Cambridge: Cambridge University Press.

Kahneman, D. & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist*, *64*(6), 515–526.

Knight, C. (2023). *Reflective Equilibrium* (Winter 2023 ed.). Metaphysics Research Lab, Stanford University. [Online; accessed 2024-01-07].

Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., & Damasio, A. (2007). Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature; London*, *446*(7138), 908–11.

Königs, P. (2020). Experimental ethics, intuitions, and morally irrelevant factors. *Philosophical Studies*, *forthcoming*, 1–19.

Kozhevnikov, M. & Hegarty, M. (2001). Impetus beliefs as default heuristics: Dissociation between explicit and implicit knowledge about motion. *Psychonomic Bulletin & Review*, *8*(3), 439–453.

Kumar, V. (2016). The empirical identity of moral judgment. *The Philosophical Quarterly*, *66*(265), 783–804.

Kumar, V. & Campbell, R. (2012). On the normative significance of experimental moral psychology. *Philosophical Psychology*, *25*(3), 311–330.

Link, S. W. & Tindall, A. D. (1971). Speed and accuracy in comparative judgments of line length. *Perception & Psychophysics*, *9*(3), 284–288.

Low, J., Apperly, I. A., Butterfill, S. A., & Rakoczy, H. (2016). Cognitive Architecture of Belief Reasoning in Children and Adults: A Primer on the Two-Systems Account. *Child Development Perspectives*, *10*(3), 184–9.

Moletti, G. (2000). *The Unfinished Mechanics of Giuseppe Moletti: An Edition and English Translation of His Dialogue on Mechanics, 1576, translated by W. R. Laird.* Toronto: University of Toronto Press.

Moors, A. (2014). Examining the mapping problem in dual process models. In *Dual process theories of the social mind* (pp. 20–34). Guilford.

Moors, A. & De Houwer, J. (2006). Automaticity: A Theoretical and Conceptual Analysis. *Psychological Bulletin*, *132*(2), 297–326.

Morewedge, C. K. & Kahneman, D. (2010). Associative processes in intuitive judgment. *Trends in Cognitive Sciences*, *14*(10), 435–440.

Nagel, T. (1997). *The Last Word.* Oxford: Oxford University Press.

Paxton, J. M. & Greene, J. D. (2010). Moral Reasoning: Hints and Allegations. *Topics in Cognitive Science*, *2*(3), 511–527.

Railton, P. (2014). The Affective Dog and Its Rational Tale: Intuition and Attunement. *Ethics*, *124*(4), 813–859.

Rawls, J. (1951). Outline of a Decision Procedure for Ethics. *The Philosophical Review*, *60*(2), 177–197. [Online; accessed 2023-08-30].

Rawls, J. (1999). *A Theory of Justice* (Revised edition ed.). Cambridge, Mass: Harvard University Press.

Rini, R. A. (2017). Why moral psychology is disturbing. *Philosophical Studies*, *174*(6), 1439–1458.

Scanlon, T. M. (2002). Rawls on justification. In S. Freeman (Ed.), *The Cambridge Companion to Rawls*, Cambridge Companions to Philosophy (pp. 139–167). Cambridge: Cambridge University Press.

Singer, P. (1974). Sidgwick and Reflective Equilibrium. *The Monist*, *58*(3), 490–517.

Singer, P. (2005). Ethics and Intuitions. *The Journal of Ethics*, *9*(3), 331–352.

Thomson, J. J. (1976). Killing, Letting Die, and The Trolley Problem. *The Monist*, *59*(2), 204–217.

Trémolière, B. & Bonnefon, J.-F. (2014). Efficient Kill–Save Ratios Ease Up the Cognitive Demands on Counterintuitive Moral Utilitarianism. *Personality and Social Psychology Bulletin*, *124*(3), 379–384.

Wasow, T. & Arnold, J. (2005). Intuitions in linguistic argumentation. *Lingua*, *115*(11), 1481–1496.