

Lecture 08 : Moral Psychology

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

Thursday, 7th March 2024

Contents

1	Dual Process Theories: Objections, Evidence and Significance	2
2	Ethical Implications of the Dual Process Theory	2
3	Against Reflective Equilibrium	3
3.1	Extending the Loose Reconstruction	3
3.2	What Is Reflective Equilibrium?	3
3.3	Why Is Reflective Equilibrium Significant?	4
3.4	Will Reflective Equilibrium Predictably Lead to Error?	4
4	Quick Objections to Greene’s Argument	5
4.1	Rini’s Objection	6
4.2	Rini’s Regress Objection	6
4.3	Königs’ Objection	7
4.4	Is the Argument Unacceptably Sceptical?	7
5	Evidence for Dual Process Theories	8
5.1	Aim	8
5.2	Three Studies and Their Predictions	8
5.3	What Did They Find and What Are Their Limits?	9
5.4	Next Steps	10
6	Conclusion: No End to Our Troubles	10
7	Appendix: Dual Process Theory and Auxiliary Hypotheses	10
7.1	What Is the Core Idea	11
7.2	What Are the Candidate Auxiliary Hypotheses?	11
7.3	Which Auxiliary Hypothesis to Choose?	12
	Glossary	12

1. Dual Process Theories: Objections, Evidence and Significance

A brief introduction to the three topics of this lecture.

This week's lecture builds on *Lecture 07*. Although you don't need to understand everything in that lecture, most of this week's lecture is about the loose reconstruction of Greene (2014)'s argument (see *Greene contra Ethics (Railgun Remix)* in *Lecture 07*).

This lecture covers three issues:

1. Significance and Extensions (*Against Reflective Equilibrium* (section §3), *Appendix: Dual Process Theory and Auxiliary Hypotheses* (section §7) and *Ethical Implications of the Dual Process Theory* (section §2))
2. *Quick Objections to Greene's Argument* (section §4) (none of them work, but perhaps you can do better)
3. *Evidence for Dual Process Theories* (section §5)

We will also have to consider evidence *against* the dual process theory of moral psychology, but that is a topic for next week.

2. Ethical Implications of the Dual Process Theory

The loose reconstruction of Greene's argument (see *Greene contra Ethics (Railgun Remix)* in *Lecture 07*) does not favour one type (e.g. deontological vs consequentialist) of ethical theory. But can it be extended, by means of an additional argument, to favour one type of ethical theory?

Singer (2005) and Greene suggest that Greene's argument (see *Greene contra Ethics (Railgun Remix)* in *Lecture 07*) can be used to make a positive case for (one from of) consequentialism (by eliminating all viable rivals).

Here is how Greene puts one idea for an argument:

'my point is simply that act consequentialism should get points for not chasing intuitions and that some of its competitors [...] should lose points for doing so. Note that the present argument also casts doubt on theories that, rather than chasing intuitions with codifying principles, simply allow our intuitions roam free. Likewise, it casts doubt on theories that purport to derive from first principles, but that are in fact intuition chasing—that is, theories that are actually attempts to get from first principles to the

intuitively right answers rather than attempts to get from first principles to wherever those principles happen to lead. (And, if you're like me, you suspect that this covers most, if not all, of act consequentialism's competition.)' (Greene 2014, p. 725)

Even if this ambitious further argument does not work, it appears that the conclusion of the loose reconstruction does remove some otherwise pressing objections to consequentialism.

3. Against Reflective Equilibrium

Does the conclusion of the loose reconstruction of Greene (2014)'s argument provide grounds to reject Rawls' method of reflective equilibrium (Rawls 1999)? This section introduces an argument for the claim that reflective equilibrium will reliably generate incorrect conclusions. (This section also presents a generalisation of the loose reconstruction: it now establishes a conclusion about not-justified-inferentially premises not only concerning particular moral scenarios but also concerning debatable moral principles.)

3.1. Extending the Loose Reconstruction

The loose reconstruction of Greene (2014)'s argument (see *Greene contra Ethics (Railgun Remix)* in Lecture 07) established a conclusion that is restricted to premises about particular moral scenarios. But it is possible to generalise the argument to a broader conclusion by elaborating on step 5. The result is this conclusion:

Not-justified-inferentially premises about particular moral scenarios, *and debatable principles*, cannot be used in ethical arguments where the aim is knowledge.

With this extension of the argument, we can use it to attempt to show that Rawls (1999)'s proposal about reflective equilibrium should be avoided. This is because it will reliably generate incorrect conclusions.

3.2. What Is Reflective Equilibrium?

One standard in ethics is Rawls' reflective equilibrium idea:

'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. 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.

3.3. Why Is Reflective Equilibrium Significant?

'To most moral philosophers who reason about substantive moral issues, it seems that the method of reflective equilibrium, or a process very similar to it, is the best or most fruitful method of moral inquiry. Of the known methods of inquiry, it is the one that seems most likely to lead to justified moral beliefs.' (McMahan 2013, p. 111)

(Incidentally, McMahan (2013) is an excellent source for a concise overview of reflective equilibrium, its relations to intuition; there is also a very brief discussion of a challenge from moral psychology. For further evaluations of reflective equilibrium, see Scanlon (2002) and Knight (2023).)

3.4. Will Reflective Equilibrium Predictably Lead to Error?

Consider an argument:

1. The not-justified-inferentially judgements you are inclined to make are an *indirect* consequence of fast processes (see *What Is the Role of Fast Processes In Not-Justified-Inferentially Judgements?* in Lecture 07).
2. Reflective equilibrium is therefore, in effect, a method of identifying principles which characterise how fast processes operate and generalising them. (Roughly doing for ethics what Aristotelians did for physics.)¹
3. But the fast processes are fast because they trade away accuracy to gain speed. (All broadly inferential processes face trade-offs between speed and accuracy; see *Preview: Ethics vs Physics* in Lecture 06.) Their function is to provide results that are accurate enough for mundane purposes in a limited but useful range of circumstances.

¹ In the lecture, I offer a bit more detail on this point (in the form of a dilemma).

4. We know, therefore, that the fast process will predictably be inaccurate in a range of cases. (Even though we cannot yet say much about which cases these are; see *Cognitive Miracles: When Are Fast Processes Unreliable?* in Lecture 07).
5. So while capable of producing valuable results within limits (much as broadly Aristotelian physics has plenty of applications), we know in advance that reflective equilibrium will reliably generate incorrect conclusions.²

Is this argument correct? As there are different varieties of reflective equilibrium (see Knight 2023), it would be worth checking which, if any, kinds of reflective equilibrium this argument works against. This could lead to an objection against (2) in the above argument.

Another line of objection might be that the dual-process theory of ethical cognition is not well supported by evidence after all (see *Evidence for Dual Process Theories* (section §5)). This could lead to an objection against (1) in the above argument.

A much bolder line of objection would be to argue that fast processes need not be inaccurate (see Railton (2014) for an attempt to develop this objection).³ This could lead to an objection against (3) in the above argument.

4. Quick Objections to Greene's Argument

Premises about particular moral scenarios, and about debatable principles, which are not-justified-inferentially cannot be used in ethical arguments where the aim is knowledge. So the conclusion of Greene (2014)'s argument, as loosely reconstructed in *Greene contra Ethics (Railgun Remix)* in Lecture

² Unless, that is, it were limited to familiar situations. But this would be hard to do given that we are not in a position to know which situations are unfamiliar (see *Cognitive Miracles: When Are Fast Processes Unreliable?* in Lecture 07). And it would also not be very useful. After all, we have little need for a theory covering only cases that our fast processes already provide us with expertise in dealing with. And if the idea of reflective equilibrium is just to identify principle implicit in responses due to fast processes, it should be subsumed into moral psychology rather than viewed as a method of doing ethics.

³ Note that Railton identifies intuition with the affective system \citep[pp.~826–8]{railton:2014_affective}. Railton is surely correct that affect is part of a flexible and sophisticated learning system. But evidence that feelings and emotions play at most a limited role in moral judgement (see *Moral Intuitions and Emotions: Evaluating the Evidence* in Lecture 02) indicates that Railton's identification of intuition and the affective system may not work.

07 and extended in *Against Reflective Equilibrium* (section §3). Given that this conclusion presents a problem for a variety of approaches to ethics, we should consider objections. Start with *quick* objections—those which do not require much additional knowledge or reasoning. If one of these succeeds, we will be spared from having to consider onerous objections.

4.1. Rini's Objection

I claim that the loose reconstruction of Greene (2014)'s argument, unlike debunking arguments, does not depend on premises about which factors are morally relevant. Rini (2016) makes an assertion which is incompatible with this claim:

‘To say that a particular psychological process does not track moral truth is to say that the process generates judgments which are not subjunctively sensitive to *certain* moral properties. We cannot say this without making some moral judgments ourselves’ (Rini 2016, p. 682, my emphasis).

Here Rini has in mind matters such as the controversy between Singer and Kamm, where they take contrasting positions on whether distance could ever be a morally relevant factor (see *Singer vs Kamm on Distance* in Lecture 06). Of course Rini is right about such cases.

But the loose reconstruction depends only on general claims about general limits of fast processes. It does not depend on any premises about whether any particular factor is morally relevant. (Indeed, the loose reconstruction is consistent with any reasonable premises about which factors are morally relevant.) Rini's assertion is false.

4.2. Rini's Regress Objection

Against debunking arguments, Rini offers an objection based on the idea that no such argument can succeed without triggering a regress:

‘nearly any attempt to debunk a particular moral judgment on grounds of its psychological cause risks triggering a regress, because a debunking argument must involve moral evaluation of the psychological cause—and this evaluation is itself then subject to psychological investigation and moral evaluation, and so on’ (Rini 2016, p. 676).

Although Rini's stated target is debunking arguments, we should ask: Does her line of objection apply to the loose reconstruction of Greene (2014)'s argument?

We can see that it does not because the regress objection works by attempting to raise doubts about the moral judgements the argument it is targeting. But the loose reconstruction of Greene's argument does not depend on specific moral judgements (see *Greene contra Ethics (Railgun Remix)* in Lecture 07).

4.3. Königs' Objection

Königs observes that debunking arguments

‘are dialectically useless if we assume that case-specific intuitions are, as a rule, subordinate to intuitions at a higher level of generality’ (Königs 2020, p. 2607).

Does the same apply to the loose reconstruction of Greene (2014)'s argument? Yes.

Is this an objection? No, for two reasons. First, the assumption Königs requires conflicts with a range of methods in ethics (see *Foot and Trolley Cases: Kant Was Wrong* in Lecture 06, *Singer vs Kamm on Distance* in Lecture 06, and *Thomson's Other Method of Trolley Cases* in Lecture 06.) Second, although the conclusion of the loose reconstruction concerns judgements about particular moral scenarios, this is only for simplicity. The argument can be straightforwardly generalised to ‘intuitions [that is, not-justified-inferentially judgements] at higher level of generality’.

4.4. Is the Argument Unacceptably Sceptical?

If the loose reconstruction of Greene (2014)'s argument succeeds, which ethical premises should we reject? The conclusion of the argument as stated is limited to not-justified-inferentially judgements about particular moral scenarios. However, the argument can be straightforwardly extended to a wider range of not-justified-inferentially judgements.

This suggests the following objection:

The loose reconstruction implies that we cannot use any not-justified-inferentially ethical judgements. But ethics depends on such judgements. So the loose reconstruction implies that ethics is impossible.

Such an objection might be especially appealing to proponents of Audi (2015, p. 57)'s view that ‘[i]ntuition is a resource in all of philosophy, but perhaps nowhere more than in ethics’ (p. 57).

To see that this objection fails, consider that a counterpart of it targeting physics rather than ethics would, at some point in history, appeared have

been no less correct than the actual objection is today. Since the counterpart targeting physics is clearly incorrect, it seems we should reject the objection.

Why does the objection fail? It relies on faith in contemporary philosophers' methods. But even passing acquaintance with intellectual history reveals that philosophers, like all researchers, can pursue mistakes in great depth over long periods of time. This is not scepticism—it's history.

5. Evidence for Dual Process Theories

What is the strongest evidence in favour of our stripped-down dual-process theory of moral psychology (see *A Dual Process Theory of Ethical Judgement* in Lecture 07)? Greene (2014) cites many studies. In this section we evaluate three of them, including one involving process dissociation (Conway & Gawronski 2013). (As a bonus, process dissociation also enables us to revisit the issue of whether emotion influences moral judgement.)

5.1. Aim

The first premise of our stripped-down dual process theory of moral psychology (see *A Dual Process Theory of Ethical Judgement* in Lecture 07) is probably the most controversial:

1. Ethical judgements are explained by a dual-process theory, which distinguishes faster from slower processes.

What is the strongest evidence in favour of this premise? Greene (2014) cites many studies. As always we should not take for granted that Greene's description of the studies is correct: we need to evaluate them for ourselves (see the step-by-step guide in *Moral Intuitions and Emotions: Evaluating the Evidence* in Lecture 02).

We are looking for evidence in favour of the stripped-down dual process theory together with our selected auxiliary hypothesis:

Auxiliary Assumption: Only the slow process ever flexibly and rapidly takes into account differences in the more distal outcomes of an action (see *Appendix: Dual Process Theory and Auxiliary Hypotheses* (section §7)).

5.2. Three Studies and Their Predictions

Here we will consider three of the studies Greene cites. It is important to specify which prediction each study tests (which may not be obvious from the abstract).

- Suter & Hertwig (2011) — prediction: limiting the time available to make a decision will reduce the influence of distal outcomes.
- Trémolière & Bonnefon (2014) — prediction: limiting the time available to make a decision will reduce consequentialist responses.⁴
- Conway & Gawronski (2013) — prediction: higher cognitive load will reduce the dominance of the more outcome-sensitive process.

5.3. What Did They Find and What Are Their Limits?

Suter & Hertwig (2011) is an example of a relatively simple study which provides evidence in favour of the dual process theory plus auxiliary hypothesis.

One limit of this study is that it does not involve any variation in the distal outcomes of actions. This is relevant because the auxiliary hypothesis is about how different processes are differently influenced by distal outcomes.

Although not designed with exactly this in mind, Trémolière & Bonnefon (2014) does observe responses to otherwise similar actions with different distal outcomes. However, the findings are not predicted by the dual process theory and auxiliary hypothesis.⁵

One limit of both Suter & Hertwig (2011) and Trémolière & Bonnefon (2014) is that they treat responses as either consequentialist or not. These studies are sometimes presented as comparing consequentialist with deontological responses; but this cannot be accurate because failing to respond as a consequentialist does not make you a deontologist (you may be neither).

Conway & Gawronski (2013) overcome this limit in addition to observing responses to otherwise similar actions with different distal outcomes. It is one of the strongest tests of the stripped-down dual process theory and its auxiliary hypothesis. These authors find, as predicted, that higher cognitive load reduces sensitivity to outcomes while not affecting sensitivity to moral prohibitions (such as on killing).

Conway & Gawronski (2013) are also important because they introduce pro-

⁴ As explained in the lecture recording, this study is associated with a second prediction, which the results appear to disconfirm: limiting the time available to make a decision will reduce sensitivity to outcomes.

⁵ Although the Trémolière & Bonnefon (2014)'s findings may be interpreted as disconfirming a prediction (as Gawronski & Beer 2017, p. 669 propose), it would be incautious to rely on post hoc reinterpretations of findings.

cess dissociation in moral psychology. Although difficult to understand (I attempt to explain the bare minimum you need in the lecture), this is a powerful method for testing theories.

5.4. Next Steps

Having evaluated some of the evidence in favour of the dual process theory, our next task is to consider evidence against it.

6. Conclusion: No End to Our Troubles

The dual process theory of ethical judgement appears, so far, to be well supported by evidence for it (although we have yet to consider any evidence against it). This supports the loose reconstruction of Greene (2014)'s argument (see *Greene contra Ethics (Railgun Remix)* in Lecture 07), which is also not vulnerable to any of the quick objections (at least none I could find in the literature). We may therefore have to accept the conclusion that not-justified-inferentially premises about particular moral scenarios, and about debatable ethical principles, cannot be used in ethical arguments insofar as the arguments aim to establish knowledge of their conclusions. This does not show that consequentialism is the one true ethical theory. But it does imply that we should avoid a variety of approaches to doing ethics, including Foot's, Kamm's and Rawls' method of reflective equilibrium insofar as our aim is to gain knowledge of ethical truths.

7. Appendix: Dual Process Theory and Auxiliary Hypotheses

The stripped-down dual process theory (see *A Dual Process Theory of Ethical Judgement* in Lecture 07) requires at least one auxiliary hypothesis in order to relate it to available evidence. At least six auxiliary hypotheses have been proposed. Which should we accept?

This is covered in the lecture in 2023–4. You do not need this material, but you may wish to consider it if you are writing about dual-process theories of ethical judgement.

Proponents of dual process theories of ethical judgments tend to offer complex sets of hypotheses which go far beyond the core idea of a dual process theory and are not all equally well supported by evidence (for two examples, see Green and Kumar in *A Dual Process Theory of Ethical Judgement* in Lecture 07).

Since the hypotheses that are associated with different proposals can be isolated and tested separately, combining them yields something more like a parlay bet than a theory.⁶ Better to keep them separate.

7.1. What Is the Core Idea

Just this and nothing more.

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.

Because much of the available evidence involves observations of just one response to a moral scenario (typically a verbal judgement⁷), additional, auxiliary hypotheses are needed to generate predictions linked to the currently available evidence. But which auxiliary hypothesis should we adopt?

7.2. What Are the Candidate Auxiliary Hypotheses?

At least six candidates for auxiliary hypotheses can be found in the existing literature (and there are probably many more):

id	fast	slow	source
1.	never consequentialist ⁸	sometimes consequentialist	Greene (2014)

⁶ This is too quick: all theories are bets. And I don't have anything against parlay bets—if you are given good enough odds, they can be spectacular (and fun, of course). But making progress in understanding is like gambling in that it requires managing risk. Just as you have to allocate cash to bets carefully in order to minimize the risk of ruin, so in your research you have to allocate time carefully in order to avoid being left with nothing. (There is some research on how this is done that starts with the Kelly Criterion.) You should not put everything into a small number of especially risky bets. This is the true problem with Parlay bets.

⁷ Measures like response time, movement trajectory, proactive eye gaze or pupil dilation do not seem to have caught on in moral psychology despite being used to good effect in supporting dual process theories in other domains including social cognition (see van der Wel et al. 2014 or Edwards & Low 2017 for some examples).

⁸ In this context, the term 'consequentialist' is short for characteristically consequentialist.

id	fast	slow	source
2.	always deontological ⁹	sometimes not deontological	Greene (2014) ¹⁰
3.	always only act-types	sometimes not only act-types	Cushman (2013)
4.	never distal outcomes	sometimes distal outcomes	Cushman (2013) ¹¹
5.	model-free	model-based	Cushman (2013)
6.	affective	cognitive	Greene (2014)

UPDATE A further candidate auxiliary hypothesis might concern the distinction between personal and impersonal. Since this introduces even more problems than we will eventually have to face (McGuire et al. 2009), it does not appear to be a good bet as things stand.

7.3. Which Auxiliary Hypothesis to Choose?

I propose that we use just one auxiliary hypothesis, which is a version of (4) above:

Only the slow process ever flexibly takes into account the more distal outcomes of an action.

This auxiliary hypothesis has two virtues. Together with the stripped-down dual process theory, it generates nearly all the predictions for which we currently have good evidence. And it cannot be significantly weakened while still doing this.

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

⁹ In this context, the term ‘deontological’ is short for characteristically deontological.

¹⁰ Greene (2014) does not explicitly separate this and the first candidate, (1). But they are clearly distinct.

¹¹ Cushman (2013) phrases the distinction in terms of actions and outcomes. This is potentially confusing if, like Davidson (1971), you think that actions are individuated by their outcomes. But it is clear from the discussion that Cushman’s distinction concerns a contrast between more proximal outcomes and more distal outcomes. Note also that Cushman does not explicitly distinguish this and the third candidate, (3); perhaps the distinction is too minor to consider.

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 14

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.’ 11

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.’ 12

cognitively efficient A process is *cognitively efficient* to the degree that it does not consume working memory and other scarce cognitive resources. 14

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). 6, 7

distal outcome The outcomes of an action can be partially ordered by the cause-effect relation. For one outcome to be more *distal* than another is for it to be lower with respect to that partial ordering. To illustrate, if you kick a ball through a window, the window’s breaking is a more distal outcome than the kicking. 8, 9, 12

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. 5, 8

fast A *fast* process is one that is 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 authors 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. 4, 6, 8, 11, 15

loose reconstruction (of an argument). A reconstruction which prioritises finding a correct argument for a significant conclusion over faithfully representing the argument being reconstructed. 2, 5, 6

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, 4, 7, 10

outcome An outcome of an action is a possible or actual state of affairs. 13, 14

proximal outcome The outcomes of an action can be partially ordered by the cause-effect relation. For one outcome to be more *proximal* than another is for it to be higher with respect to that partial ordering. To illustrate, if you kick a ball through a window, the kicking is a more proximal outcome than the window’s breaking. 12

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, 4

slow converse of fast. 8

unfamiliar problem An unfamiliar problem (or situation) is one ‘with which we have inadequate evolutionary, cultural, or personal experience’ (Greene 2014, p. 714). 5

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.
- Conway, P. & Gawronski, B. (2013). Deontological and utilitarian inclinations in moral decision making: A process dissociation approach. *Journal of personality and social psychology*, 104(2), 216–235.
- Cushman, F. (2013). Action, Outcome, and Value: A Dual-System Framework for Morality. *Personality and Social Psychology Review*, 17(3), 273–292.
- Davidson, D. (1971). Agency. In R. Binkley, R. Bronaugh, & A. Marras (Eds.), *Agent, Action, and Reason*, (pp. 3–25). Toronto: University of Toronto Press. Reprinted in Davidson, D. (1980) *Essays on Actions and Events*. Oxford: Oxford University Press.
- Edwards, K. & Low, J. (2017). Reaction time profiles of adults’ action prediction reveal two mindreading systems. *Cognition*, 160, 1–16.
- 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. & Beer, J. S. (2017). What makes moral dilemma judgments “utilitarian” or “deontological”? *Social Neuroscience*, 12(6), 626–632.

- 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.
- Kahneman, D. (2002). Maps of bounded rationality: A perspective on intuitive judgment and choice. In T. Frangmyr (Ed.), *Le Prix Nobel, ed. T. Frangmyr, 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. & 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].
- Königs, P. (2020). Experimental ethics, intuitions, and morally irrelevant factors. *Philosophical Studies, forthcoming*, 1–19.
- McGuire, J., Langdon, R., Coltheart, M., & Mackenzie, C. (2009). A reanalysis of the personal/impersonal distinction in moral psychology research. *Journal of Experimental Social Psychology, 45*(3), 577–580.
- McMahan, J. (2013). *Moral Intuition*, (pp. 103–120). Wiley. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/b.9780631201199.1999.00007.x>.
- 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.
- 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. (2016). Debunking debunking: A regress challenge for psychological threats to moral judgment. *Philosophical Studies*, 173(3), 675–697.
- 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.
- Suter, R. S. & Hertwig, R. (2011). Time and moral judgment. *Cognition*, 119(3), 454–458.
- 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.
- van der Wel, R. P. R. D., Sebanz, N., & Knoblich, G. (2014). Do people automatically track others' beliefs? evidence from a continuous measure. *Cognition*, 130(1), 128–133.