Lecture 03 : Ethics : Two Systems

Corrado Sinigaglia & Stephen A. Butterfill

< >

Monday, 28th March 2022

Contents

1. Ethical Cognition

1.1. Why This Two-Systems Theory?

Greene offers an elaborate dual-process theory of ethical cognition, one which incorporates controversial claims about consequentialism and emotion.[1] As these claims are neither essential features of a dual-process theory nor necessary for the overall argument we are developing, we may consider a stripped-down dual process theory instead.

1.2. The Stripped-Down Two-Systems 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 two-systems 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.

1.3. What Does the Two-systems Theory Predict?

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

> Only the slow process ever flexibly and rapidly takes into account differences in the more distal outcomes of an action.

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

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

---

[1] 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).

'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).[2]

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.

We will consider some more evidence for, and against, the dual-process theory.

1.4. Other Two-Systems Theories of Ethical Cognition

Can be found in the notes to another course:

- A Dual Process Theory of Ethical Judgement

1.5. Evidence For

What is the strongest evidence in favour of the stripped-down two-systems theory of ethical cognition? Greene (2014) cites many studies. Here we will consider three of them:

- 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 characteristically consequentialist responses.[3]

- Conway & Gawronski (2013) — prediction: higher cognitive load will reduce the dominance of the slow process and threfore reduce the influence of distal outcomes.

1.5.1. What Did These Three Studies Find?

Suter & Hertwig (2011) is an example of a relatively simply study which pro-

---

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

[3] 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.

vides 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.[4]

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 would 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 process dissociation in moral psychology. Although difficult to understand, this is a powerful method for testing two-systems theories.[5]

1.6.   Are the Studies Really Evidence for Two Systems?

1.6.1.   Time Pressure

Recall that Suter & Hertwig (2011) provide evidence that time pressure makes participants less sensitive to distal outcomes. Bago & Neys (2019) consider what happens when subjects first make a moral judgement under time pressure and extraneous cognitive load and then, just after, make another moral judgement (in answer to the same question) with no time pressure and no extraneous cognitive load. They report:

> 'Our critical finding is that although there were some instances
> in which deliberate correction occurred, these were the excep-

---

[4]   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.

[5]   Process dissociation has been also used in testing two-systems theories of memory (Jacoby 1991) and perspective taking (Todd et al. 2016).

tion rather than the rule. Across the studies, results consistently showed that in the vast majority of cases in which people opt for a [consequentialist] response after deliberation, the [consequentialist] response is already given in the initial phase' (Bago & Neys 2019, p. 1794).

This is an obstacle to considering Suter & Hertwig (2011)'s study as evidence for our dual-process theory of moral judgement.

1.6.2.   Process Dissociation

Recall that Conway & Gawronski (2013) use process dissociation to provide evidence for the prediction that higher cognitive load reduces sensitivity to more distal outcomes.

Gawronski et al. (2017) note that reduced sensitivity to more distal outcomes could be consequence of a general preference not to do anything when under time pressure. They therefore extend the process dissociation model to include a preference for no action.

Separating sensitivity to distal outcomes from preferences not to act changes the picture:

> 'The only significant effect in these studies was a significant increase in participants' general preference for inaction as a result of cognitive load. Cognitive load did not affect participants' sensitivity to morally relevant consequences' (Gawronski et al. 2017, p. 363).

They conclude:

> 'cognitive load influences moral dilemma judgments by enhancing the omission bias, not by reducing sensitivity to consequences in a utilitarian sense' (Gawronski et al. 2017, p. 363).

While we should be cautious about putting too much weight on this study, these results do reveal that we cannot take Conway & Gawronski (2013) as evidence in favour of our dual-process theory and auxiliary hypothesis.

1.6.3.   Conflicts in the Conflicting Evidence

The two studies which conflict with the evidence for our dual-process theory also appear to conflict with each other. If Gawronski et al. (2017) is right about cognitive load, the participants in Bago & Neys (2019)'s study should have appeared to be less 'utilitarian' (as they describe it) when under cognitive load. This is because avoiding action would lead one to make judgements that Bago & Neys classify as non-utilitarian.

So we cannot accept both Gawronski et al. (2017)'s and Bago & Neys (2019)'s conclusions.

This is a sign that there may be something wrong with the way the studies are constructed, perhaps because the dual-process theories they are targeting are not well specified (e.g. involve too many independent bets being made simultaneously).

1.7.   Conclusion

We may not yet have found sufficient grounds to accept or reject the stripped-down two-systems of ethical cognition.

2.   Ethics: Significance of Two Systems

2.1.   Argument Outline

What is the two systems theory of ethical cognition significant?   Because it conflicts with the widespread (in philosophy) use of not-justified-inferentially premises in arguments intended to provide knowledge of the truth of their conclusions.

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

2. Faster processes are unreliable in unfamiliar situations.

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.

5. We have reason to suspect that the moral scenarios and principles philosophers consider involve unfamiliar situations.

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

2.2.   Case Study: Thomson's Method of Trolley Cases

To see why the conclusion of the argument above is significant, we need to see how many philosophers approach ethics.

Consider Thomson (1976) on what she calls 'the 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? I like to call this the trolley problem, in honor of Mrs. Foot's example' (Thomson 1976, p. 206).

Foot (1967) had earlier suggested that it is at least in part because duties not to harm rank above duties to help. To counter this suggestion, Thomson adds a further trolley case:

> '**Frank** is a passenger on a trolley whose driver has just shouted that the trolley's brakes have failed, and who then died of the shock. On the track ahead 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 Frank can turn the trolley onto it. Unfortunately there is one person on the right-hand track. Frank can turn the trolley, killing the one; or he can refrain from turning the trolley, letting the five die' (Thomson 1976, p. 207).

Frank's case is constructed in such a way that (according to Thomson[6]) if he does nothing, he fails to help; whereas if turns the trolley, he harms one person in order to help five. His choice is between harming one or helping five. Thomson infers:

> 'By her [Foot's] principles, Frank may no more turn that trolley than David may cut up his healthy specimen' (Thomson 1976, p. 207).[7]

Thomson responds by relying on what appears to be an empirical claim:

> 'Yet I take it that anyone who thinks Edward may turn his trolley will also think that Frank may turn his' (Thomson 1976, p. 207).

It is possible to interpret Thomson as offering this as a normative claim (anyone *must* take it to be so). Alternatively, she might consider her position as one that is relevant only to those who agree with her on this. So there is no obvious commitment to an empirical claim here.

---

[6] This qualification is necessary because there is a tricky issue about which, if any, omissions are actions. If Frank's refraining from turning the trolley is an action which harms the five, then Frank's choice is between harming one and harming five and so his case does not work against Foot in the way Thomson intends.

[7] Here Thomson appears to misrepresent Foot's position. Foot (1967, p. 17) stresses, 'I have not, of course, argued that there are no other principles.' But the key issue is not whether Foot is right but whether the principle that duties not to harm rank above duties to help can justify the pattern of judgements.

In any case, Thomson takes the pattern of judgements about what David, Edward and Frank should do to justify rejecting Foot's view[8] in favour of her own:

> 'what matters in these cases in which a threat is to be distributed is whether the agent distributes it by doing something to it, or whether he distributes it by doing something to a person' (Thomson 1976, p. 216).

If the above loose reconstruction of Greene's argument is correct, Thomson's method of trolley cases is misguided because it relies on not-justified-inferentially premises about particular moral scenarios.

2.3. Further Implication

The loose reconstruction of Greene's 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).[9] (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, 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)

---

[8] Note that Thomson is rejecting only Foot's answer to the trolley problem. Thomson (1976, p. 217) concedes, 'Mrs. Foot and others may be right to say that negative duties are more stringent than positive duties.'

[9] 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.

use an Aristotelian theory of the physical to characterise physical cognition). Or the aim might be to understand what consistency with certain judgements would require.

2.4.   Generalisation to Other Domains

Can the loose reconstruction of Greene's argument concerning ethics be generalised to other domains? On the face of it, none of the arguments for the premises rely on features are specific to ethics.

2.5.   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[10]).

Why bother with my loose reconstruction when we 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.

Glossary

automatic On this course, 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,

---

[10] 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.

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 11

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.' 2, 3

cognitively efficient A process is *cognitively efficient* to the degree that it does not consume working memory and other scarce cognitive resources. 11

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). 7, 8, 12

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). 9

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. 2–5

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. 2

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). 7, 8, 12

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. 2, 6, 12

Frank 'Frank is a passenger on a trolley whose driver has just shouted that the trolley's brakes have failed, and who then died of the shock. On the track ahead 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 Frank can turn the trolley onto it. Unfortunately there is one person on the right-hand track. Frank can turn the trolley, killing the one; or he can refrain from turning the trolley, letting the five die' (Thomson 1976, p. 207). 7, 8

loose reconstruction (of an argument). A reconstruction which prioritises finding a correct argument for a significant conclusion over faithfully representing the argument being reconstructed. 9

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. 6, 8

outcome An outcome of an action is a possible or actual state of affairs. 10

slow converse of fast. 3, 6

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). 8

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

References

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

Bago, B. & Neys, W. D. (2019). The Intuitive Greater Good: Testing the Corrective Dual Process Model of Moral Cognition. *Journal of Experimental Psychology: General*, *148*(10), 1782–1801.

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.

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., 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.

Foot, P. (1967). The problem of abortion and the doctrine of the double effect. *Oxford Review*, *5*, 5–15.

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.

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.

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., 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.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*(5), 513–541.

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

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

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.

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

Suter, R. S. & Hertwig, R. (2011). Time and moral judgment. *Cognition*, *119*(3), 454–458.

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

Todd, A. R., Cameron, C. D., & Simpson, A. (2016). Dissociating processes underlying level-1 visual perspective taking in adults. *Cognition*, *in press.*

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.