Lecture 03 : Moral Psychology

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

Thursday, 25th January 2024

Contents

1. Introduction to Lecture 3: Emotion and Reason in Moral Judgement

How, if at all, does a person's reasoning influence their moral judgements? This is our next question. Until now we were focussed on moral intuitions and emotion. Now it is time to consider reasoning. We will do this by critically reviewing research on two key phenomena which illustrate roles for reasoning: moral dumbfounding and moral disengagement.

How, if at all, does a person's reasoning influence their moral judgements?

Several authors have defended, on the basis of evidence we shall consider in this lecture, strong views:

> moral 'judgments are [not] the conclusions of explicitly represented syllogisms, one or more premises of which are moral principles, that ordinary folk can articulate.' (Dwyer 2009, p. 294)

> 'If we ask people why they hold a particular moral view [their] reasons are often superficial and post hoc. If the reasons are successfully challenged, the moral judgment often remains [...] basic values are implemented in our psychology in a way that puts them outside certain practices of justification' (Prinz 2007, p. 32).

> 'moral reasoning is [...] usually engaged in after a moral judgment is made, in which a person searches for arguments that will support an already-made judgment' (Haidt & Bjorklund 2008, p. 189).

Since Dwyer and Prinz both infer these claims from a phenomenon called moral dumbfounding (and Haidt is one of the authors of the first report on dumbfounding), we will start our investigation with dumbfounding.

1.1. Illustration: Tamanda and Jonty

Tamanda and Jonty are both refereeing a school match between a privileged team that always wins and a team of underdogs that has never won in the entire history of the game. To the privileged team, win or lose will make little difference in the long run. But if the underdogs lose, their team will be disbanded and future generations will not get to train and play at all. The underdogs are about to win when Tamanda and Jonty must each independently decide whether the underdogs have fouled. They are each quite confident, but not very confident, that there was a foul. Not calling a foul risks unfairness; calling a foul risks harming the underdog's future for decades to come.

Tamanda and Jonty are now each faced with an ethical decision: in deciding whether they have sufficient evidence to call a foul, may they require an unusually high level of confidence on the grounds that so much is at stake for the underdogs?

Tamanda and Jonty have different approaches to making ethical decisions. Jonty does what he feels is right. If asked to give reasons, he will provide a justification; but in his case the justification is always entirely constructed after the decision.

Tamanda takes a different approach. She has been reading Scanlon (1998), a philosophical theory in ethics, and is attempting to live out these ideas as closely as possible in her everyday life. Tamanda's core conviction is that she must follow principles that no one can reasonably reject. For each of the actions she may choose—calling a foul and not calling a foul—she therefore attempts to work out whether it would be disallowed by principles that no one could reasonably reject.[1] Her aim is never to act in ways that violate such principles.

Tamanda and Jonty are idealisations, not people who live and breath among us, of course. But they illustrate two extreme views on how a person's reasoning might influence (or not) their moral judgements. Versions of both extremes can be found in the literature, as well as arguments for intermediate positions (Hindriks 2014).

1.2. Related Philosophical Issues

This section is not part of the spoken lecture.

If we ask narrowly philosophical questions about moral justification, or about how moral knowledge is possible in principle, or about what distinguishes the moral from the non-moral, we can easily identify views on which reasoning is everything and emotions or feelings play no role (Scanlon 1998, for example). And we can identify philosophers who have taken a converse view (Nichols 2004, for example).

Those are not our questions.

Our question is about actual people's ethical abilities: we seek to understand how, if at all, reasoning influences people's moral judgements.

It is important to keep these questions separate. It could turn out, for example, that reasoning plays no significant role in how people arrive at moral

[1] Compare Scanlon (1998, p. 153): 'act is wrong if its performance under the circumstances would be disallowed by any set of principles for the general regulation of behavior that no one could reasonably reject as a basis for informed, unforced general agreement.'

judgements while also being true—perhaps tragically—that reasoning alone is the source of moral knowledge.

Nevertheless, there may be ways in which the narrowly philosophical questions and our question relate (philosophers who have attempted to make connections include Nichols 2004 and Hindriks & Sauer 2020).

2. Moral Dumbfounding

Moral dumbfounding is 'the stubborn and puzzled maintenance of an [ethical] judgment without supporting reasons' (Haidt et al. 2000, p. 1). By the end of this section you should know what moral dumbfounding is and be familiar with some of the scientific research taken to establish that, and question whether, it occurs.

Moral dumbfounding is 'the stubborn and puzzled maintenance of a judgment without supporting reasons' (Haidt et al. 2000, p. 1).

The most cited evidence for dumbfounding comes from some unpublished (!) research which is presented in the recording (Haidt et al. 2000). This research hinges on two contrasts:

1. morally provocative but harmless events vs nonmorally provocative but harmless events; and
2. morally provocative events that are harmless vs morally provocative scenarios involving harm

Examples of morally provocative but harmless events:

> '[Incest] depicts consensual incest between two adult siblings, and […] [Cannibal] depicts a woman cooking and eating a piece of flesh from a human cadaver donated for research to the medical school pathology lab at which she works. These stories were … were carefully written to be harmless' (Haidt et al. 2000).

The other scenarios commonly used in studies of moral dumbfounding are Heinz and Trolley.

2.1. An Effect of Cognitive Load?

'In Study 2 [which is not reported in the draft] we repeated the basic design while exposing half of the subjects to a cognitive load—an attention task that took up some of their conscious mental work space—and found that this load increased the level of moral dumbfounding without changing subjects' judgments or their level of persuadability' (Haidt & Bjorklund 2008, p. 198).

Further evidence for an effect of cognitive load is provided by McHugh et al. (2023).[2]

2.2. Can We Rely on Haidt et al. (2000) as Evidence?

Before relying on any study we must check whether there are (i) successful or unsuccessful replications,) (ii) similar studies with convergent or divergent results, and (iii) reviews or metaanalyses in which the study features (see *Moral Intuitions and Emotions: Evaluating the Evidence* in Lecture 02).

Royzman et al. (2015) claim to have unsuccessfully replicated the unpublished research on moral dumbfounding:

> '3 of [...] 14 individuals [without supporting reasons] disapproved of the siblings having sex and only 1 of 3 (1.9%) maintained his disapproval in the "stubborn and puzzled" manner' (Royzman et al. 2015, p. 309).

They conclude that:

> 'a definitionally pristine bout of MD is likely to be a extraordinarily rare find, one featuring a person who doggedly and decisively condemns the very same act that she has no prior normative reasons to dislike' (Royzman et al. 2015, p. 311).

But your lecturer is unconvinced by this. They did, in fact, find one person who was dumbfounded even by their own criteria. Further, Haidt et al. (2000)'s method is to compare morally provocative events that are harmless with morally provocative scenarios involving harm.[3] Their prediction is that their should be *significantly more* dumbfounding in the former. Royzman et al. (2015) have not designed an experiment which tests this prediction.

McHugh et al. (2017) offers a successful replication.[4] These results were extended in McHugh et al. (2020) and McHugh et al. (2023).

---

[2] McHugh et al. (2023) do make a strong case for the effect of cognitive load on reducing reasoning generally. But note that these researchers did not find evidence either way concerned effects of cognitive load in the *Incest* scenario. They speculate that this could be due to lack of statistical power.

[3] Compare Haidt et al. (2000): 'They made the fewest such declarations in Heinz, and they made significantly more such declarations in the Incest story.'

[4] Note that in McHugh et al. (2017), Study 1 is a bit different from the other studies. In Study 1, there is a robust distinction between the 'reasoning' dilemmas (Heinz and Trolley) and the 'intuition' dilemmas (Incest and Cannibal). In Study 3a and 3b together there is weaker evidence for this distinction. But Studies 2+ were all online studies, and I am not persuaded that they actually worked (perhaps participants were simply rushing through the questions?).

McHugh et al. (2023) investigated moral dumbfounding with participants drawn from three different regions: China, India and North Africa and the Middle East. They found evidence for moral dumbfounding in all cases, with variation in which dilemmas invoked most dumbfounding:

> 'for both the Indian sample and the MENA sample, Trolley appeared to evoke the highest rates of dumbfounding, while for the Chinese sample, Cannibal evoked the highest rates of dumbfounding. In contrast for WEIRD samples, Incest tends to be the scenario that most reliably evokes dumbfounding (McHugh et al. 2023, p. 1056)

2.3. Do We Even Need a Study?

It seems quite easy to elicit moral dumbfounding in everyday life. This is something you could try for yourself.[5]

2.4. Appendix: Philosophical Perspectives

This section is not in the spoken lecture

Two recent discussions of dumbfounding are Guglielmo (2018) and Wylie (2021).

3. Why Is Moral Dumbfounding Significant?

Moral dumbfounding is used in a variety of philosophical arguments. Dwyer (2009) argues that moral dumbfounding provides evidence for what she calls 'The Linguistic Analogy'. Prinz (2007) argues that moral dumbfounding supports the view that emotions alone, not reasoning, determines which moral judgements humans make. This section critically evaluates both arguments. Have their proponents understood moral dumbfounding?

3.1. What Does Moral Dumbfounding Show? A Misconstrual

Dwyer (2009, p. 294) takes the evidence for moral dumbfounding to show that

> moral 'judgments are [not] the conclusions of explicitly represented syllogisms, one or more premises of which are moral principles, that ordinary folk can articulate.'

[5] Be careful; it turns out that some people react badly if you ask them about incest and eating their pets.

This is a mistake. The abstract for Haidt et al. (2000) states:

> 'It was hypothesized that participants' judgments would be highly consistent with their reasoning on the moral reasoning dilemma' [ie. reasoning concerning the morally provocative and harmfull events].

And this is what those researchers found.

Moral dumbfounding is investigated a matter of degree: some dilemmas lead to greater moral dumbfounding than others (and which dilemmas lead to more dumbfounding varies from place to place, as McHugh et al. (2023) show).[6]

3.2. What Does Moral Dumbfounding Truly Show?

The existence of moral dumbfounding shows that *some* moral intuitions (and thus some moral judgements) are not consequences of reasoning from known principles.

The existence of moral dumbfounding does not show that *no* moral judgements are consequences of reasoning from known principles. Indeed, reflection on moral disengagement suggests that this is false.

4. Reason and Atrocity: Hindriks' Observation

Moral reasoning appears to enable humans to condone and commit atrocities. Yet it is quite widely held that reasoning is 'usually engaged in after a moral judgment is made' (Haidt & Bjorklund 2008, p. 189). Hindriks observes (in effect) that it is hard to see how both views could be correct (Hindriks 2014; Hindriks 2015).

One compelling reason for studying moral psychology is that ethical abilities appear to play a central role in atrocities:

> 'The massive threats to human welfare stem mainly from deliberate acts of principle, rather than from unrestrained acts of impulse' (Bandura 2002, p. 116).

Further, the principles that underpin humans' capacities to perform inhumane acts are often appear to be manufactured and maintained through rea-

[6] Not everyone would agree. Royzman et al. (2015) instead present a set of criteria which must be met for a 'definitionally pristine bout of' moral dumbfounding. But as they do not find evidence for such things, it is unclear why we should abandon Haidt et al. (2000)'s approach of comparing dilemmas to find varying degrees of moral dumbfounding.

soning to fit a particular situation.[7]

This observation appears to be in tension with views on which reason can play only an indirect role in motivating morally-relevant actions (for example, harming or helping another person).

As one example of a view on the limits of reason, consider Prinz. Commenting on moral dumbfounding, Prinz (2007, p. 29) writes:

> 'If we ask people why they hold a particular moral view, they may offer some reasons, but those reasons are often superficial and post hoc. If the reasons are successfully challenged, the moral judgment often remains. When pressed, people's deepest moral values are based not on decisive arguments that they discovered while pondering moral questions, but on deeply inculcated sentiments.'

From this Prinz draws a bold conclusion:

> 'basic values are implemented in our psychology in a way that puts them outside certain practices of justification. Basic values provide reasons, but they are not based on reasons. … basic values seem to be implemented in an emotional way' (Prinz 2007, p. 32).

Prinz appears to be ignoring a key feature of the experiment he is discussing: it is structured as a comparison between harmless and harm-involving cases where subjects' level of dumbfounding differs between these (see *Moral Dumbfounding* (section §2)). The evidence he is (misre)presenting in favour of it actually challenges his view.

Haidt & Bjorklund articulate a slightly less radical view:

> 'moral reasoning is an effortful process (as opposed to an automatic process), usually engaged in after a moral judgment is made, in which a person searches for arguments that will support an already-made judgment' (Haidt & Bjorklund 2008,

[7] To take just one example, Osofsky et al. (2005) investigated prison workers who were tasked with work related to executions. They observe

> 'The executioners, who face the most daunting moral dilemma, made the heaviest use of all of the mechanisms for disengaging moral self-sanctions. They adopted moral, economic, and societal security justifications for the death penalty, ascribed subhuman qualities to condemned inmates, and disavowed a sense of personal agency in the taking of life' (Osofsky et al. 2005, p. 387).

p. 189).[8]

Hindriks observes (in effect) that even this less radical view appears to conflict with the idea that moral reasoning often appears to be necessary for condoning and performing inhumane acts (Hindriks 2014; Hindriks 2015). Affective support for judgements about not harming can be overcome with reason. Affective obstacles to deliberately harming other people can be overcome with reason. This should not be possible if reason usually occurs after a moral judgement is made and enables people only to provide post hoc justification for it.[9]

So is moral reasoning 'usually engaged in after a moral judgment is made'? Or is it essential for overcoming affective support for judgements about not harming? This discussion can be sharpened by considering moral disengagement.

5. Moral Disengagement: The Theory

Moral disengagement occurs when self-sanctions are disengaged from inhumane conduct. It enables people to do wrong and feel good.

To understand moral disengagement, we need to consider the theory, the evidence which supports it and its significance for understanding humans' ethical abilities. Start with the theory.

Bandura, who is responsible for introducing the notion of moral disengagement, offers a conjecture about self-regulation:

> 'individuals adopt standards of right and wrong [and they] monitor their conduct and the conditions under which it occurs, judge it in relation to their moral standards and perceived circumstances, and regulate their actions by the consequences they apply to themselves' (Bandura 2002, p. 102).

---

[8] This is only half of those authors' view about reasoning. They also claim that 'Moral discussion is a kind of distributed reasoning, and moral claims and justifications have important effects on individuals and societies' (Haidt & Bjorklund 2008, p. 181). Their idea, very roughly, is that moral discussion can have a long-term effect on affect which can in turn modulate individuals' judgements and actions.

[9] Hindriks focuses on a normative question about justification for moral judgements. The fact that Bandura and other social scientists tend to study abysmal bits of moral reasoning (e.g. 'Kids who get mistreated usually do things that deserve it' (Bandura et al. 1996)) is therefore a potential problem he needs to resolve (Hindriks 2014, p. 205). We need not consider this problem because our primary concern is to only understand the causal role of reason in how moral judgements are acquired.

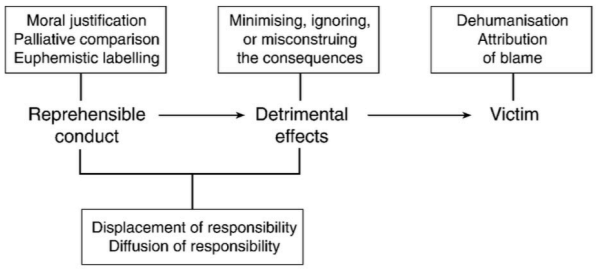

Figure 1: A model of eight mechanisms of moral disengagement *Source*:Bandura (2002, figure 1)

The upshot of self-regulation is that people constrain themselves not to violate their own moral standards. (As Bandura puts it, 'It is through the ongoing exercise of evaluative self-influence that moral conduct is motivated and regulated' (Bandura 2002, p. 102).)

This self-regulation will sometimes prevent people from getting things they want. But people can anticipate the effects of self-regulation and work around them:

> 'In the face of situational inducements to behave in inhumane ways, people can choose to behave otherwise by exerting self-influence' (Bandura 2002, p. 102).

This 'self-influence' amounts to construing actions which would otherwise be incompatible with an individual's standards of right and wrong in ways that avoid the incompatibility. In effect, self-regulation is anticipatorily derailed.

This is *moral disengagement*: the derailing of self-regulation to allow actions which would violate one's own standards of right and wrong.

Bandura postulates eight processes by which moral disengagement can occur:

> 'The disengagement may centre on redefining harmful conduct as honourable by moral justification, exonerating social comparison and sanitising language. It may focus on agency of action so that perpetrators can minimise their role in causing harm by diffusion and displacement of responsibility. It may involve minimising or distorting the harm that follows from detrimental actions; and the disengagement may include dehumanising and blaming the victims of the maltreatment' Bandura (2002, p. 103).

Their operation is depicted in the figure:

Reason plays a role in most, if not all, of these processes. It is central to Moral Justification, Displacement of Responsibility and Attribution of Blame. So if moral disengagement is responsible for a moral judgement or action, it is likely that reasoning will have played a causal role in arriving at the judgement or action.

What evidence motivates accepting Bandura's theory?

6. Moral Disengagement: The Evidence

A variety of evidence indicates that moral disengagement is a valid and useful construct.

Having understood the theory, we now need to ask, What evidence supports the view that moral disengagement occurs? And is there evidence that it can explain morally-relevant judgements and actions?

Bandura et al. (1996) constructed a questionnaire with four items for each of the eight postulated mechanisms. To illustrate with just one of the four items (the questionnaire was used to study bullying in 10–15 year old children):

1. Moral justification - 'It is alright to lie to keep your friends out of trouble.'
2. Euphemistic language - 'Slapping and shoving someone is just a way of joking.'
3. Advantageous comparison - 'It is okay to insult a classmate because beating him/her is worse.'
4. Displacement of responsibility - 'If kids are living under bad conditions they cannot be blamed for behaving aggressively.'
5. Diffusion of responsibility - 'If a group decides together to do something harmful it is unfair to blame any kid in the group for it.'
6. Distorting consequences - 'Children do not mind being teased because it shows interest in them.'
7. Attribution of blame - 'If people are careless where they leave their things it is their own fault if they get stolen.'
8. Dehumanization - 'Some people have to be treated roughly because they lack feelings that can be hurt.'

The results indicates that a single factor could be regarded as responsible for subjects' responses on all items.[10] This factor correlated significantly with antisocial behaviour, among other things. Those who scored highly on this factor

> 'tend to be more irascible, ruminate about perceived grievances, and are neither much troubled by guilt nor feel the need to make amends for harmful conduct. They also engage in a higher level of interpersonal aggression and delinquent behavior' (Bandura et al. 1996, p. 368).

This indicates that the theory of moral disengagement may be correct (or at least useful), and that the questionnaire measures moral disengagement.

Further support for these conclusions is provided by a study using the questionnaire with a demographically different population (single-parent African Americans, vs Italians), which replicated key findings (e.g. the single factor) and generated broadly congruent results overall (Pelton et al. 2004).

The measure of moral disengagement did not correlate with socioeconomic factors in either study (Bandura et al. 1996, p. 371; Pelton et al. 2004, p. 38).[11] This is important because any such correlation would not be explained by the theory of moral disengagement and could indicate that the questionnaire fails to capture a useful construct.

Variants of scale have also been developed and found useful. For example, Boardley & Kavussanu (2007) provide evidence that antisocial behaviours in sport are linked to moral disengagement. Osofsky et al. (2005) found that moral disengagement plays a role in enabling prison workers to perform tasks essential for executing prisoners. And McAlister et al. (2006) compared moral disengagement in the United States before and after the September 11th terrorist strike, finding a significant increase in moral disengagement which was correlated with a significant increase in support for the use of military force. Strikingly, these authors found that the terrorist strike itself appeared to have no effect on support for the use of military force other than through increased moral disengagement (p. 156).

Overall, we have sufficient grounds to accept that moral disengagement occurs, and that it can explain some morally-relevant judgements and actions.

But why is moral disengagement relevant to our concerns with moral psy-

[10] See Bandura et al. (1996, p. 367): 'A principal-components factor analysis with varimax orthogonal rotation revealed a single factor structure.'

[11] McAlister et al. (2006, pp. 151–2, 160), who used an 11-item questionnaire with a U.S. adult population do report effects on moral disengagement of education, ethnicity, age and location.

chology?

7. Moral Disengagement: Significance

The existence of moral disengagement shows that some moral judgements are, at least in part, consequences of reasoning from known principles. It also appears to be a source of objections to each of the theories of moral judgements we have so far considered, as well as (to anticipate) to Greene's dual-process theory.

We have understood the theory of moral disengagement and seen evidence that it occurs and can explain an interesting range of morally-relevant judgements and actions. No doubt, then, that it is interesting for its own sake. But why are we focussing on it at this point in the course on moral psychology?

The existence of moral disengagement shows that some moral judgements are, at least in part, consequences of reasoning from principles which the reasoner can articulate.[12]

The role of reason in moral disengagement—and therefore in moral judgement—is incompatible with views on which 'basic values are implemented in our psychology in a way that puts them outside certain practices of justification' (Prinz 2007, p. 32). It is also incompatible with the view that 'moral reasoning is [...] usually engaged in after a moral judgment is made, in which a person searches for arguments that will support an already-made judgment' (Haidt & Bjorklund 2008, p. 189).[13],[14]

How might a defender of one of those views rely to this objection? Observe that those views are not incompatible with the obvious truth that philoso-

[12] Royzman et al. (2014) provide an independent source of evidence for this conclusion. (Why not use this as a shortcut rather than discussing the more complicated research on moral disengagement? Because, as noted below, there are some further conclusions that we can draw by from the existence of moral disengagement.)

[13] Dahl & Waltzer (2018, p. 241) offer a conflicting interpretation: according to them, the findings about moral disengagement are 'consistent with recent proposals that decisions about moral issues do not typically follow from reasoning about moral principles [...] Instead, decisions are said to happen before moral reasoning in most situations. [...] moral reasoning happens primarily when people later seek to justify their decisions to themselves or others.' I reject their interpretation because do not know how to reconcile it with Bandura (2002, p. 102)'s point that moral disengagement requires *anticipating* the effects self-regulation; this appears to require reasoning in order to make or sustain a moral judgement.

[14] Much of research on moral disengagement does appear to support these authors' claims about the social role of reason. But note that these are independent claims. We can consistently hold that moral reasoning influences moral judgements both intra- and inter-individually.

phers sometimes reason about ethical dilemmas. This does not contradict those views because such reasoning is (presumably) relatively rare and has not been a significant part to everyday life for most people over the last hundered thousand years or so. The defender might reply to the objection by asserting that moral disengagement is similarly rare.

Is this reply correct? Moral disengagement is implicated in a wide range of inhumane actions, from small-scale bullying (Pelton et al. 2004) through executions of individuals (Osofsky et al. 2005) to the use of military force where civilian casualties are expected (McAlister et al. 2006). These findings are a challenge to the correctness of the reply.

For those who hold that reasoning does matter for moral judgement, the findings about moral disengagement may also present a challenge. The challenge is that, in moral disengagement, people's reasoning is usually terrible. Moral disengagement indicates that reasoning often functions to support moral judgements in ways that do not provide justification (because the reasoning is so bad; e.g. 'Kids who get mistreated usually do things that deserve it' (Bandura et al. 1996, p. 374)).[15] Although not directly our concern in moral psychology, this may be a source of objections to theories of moral judgements based on analogies with language (for example, Mikhail 2007).

In short, moral disengagement appears to be a potential source of objections to each of the theories of moral judgements we have so far considered.

8. Conclusion: Yet Another Puzzle

The research on dumbfounding and disengagement confronts us with a third puzzle which any acceptable theory of moral intution and action should solve.

Our overall question for this lecture was, How, if at all, does a person's reasoning influence their moral judgements?

From the two bodies of research on moral dumbfounding and moral disengagement, we can conclude that any answer to this question must be consistent the discovery that moral judgements are sometimes, but not always, a consequence of reasoning from known principles.

Further, there is no special reason to suppose that reasoning might not give rise to characteristically deontological as well as characteristically conse-

[15] Hindriks (2014, pp. 206–7) attempts to argue that individual differences in propensity to morally disengage do suggest there is a role for reason in justifying moral judgements. I think Royzman et al. (2014)'s findings would provide a more direct route to this conclusion.

quentialist moral judgements; or contractualist judgements (Scanlon 1998) or any other kind of moral judgements.

As we have seen, this is a problem for proponents of a Linguistic Analogy.[16] It is also an objection to several researchers' strong claims about reasoning functioning only for giving retrospective justification for moral judgements that have already been made (as we saw in *Moral Disengagement: Significance* (section §7)). Looking forward (hold this in mind for the future), it is also perhaps a problem for to Greene's dual-process theory, which we have yet to encounter (Greene et al. 2008; Greene 2014).

8.1. Four Puzzles (Review)

We have now seen four puzzles …

> [emotion] Why do feelings of disgust (and perhaps other emotions) sometimes influence moral judgements? And why do we sometimes feel disgust in response to moral transgressions? (see *Moral Intuitions and Emotions: Evaluating the Evidence* in Lecture 02)
>
> [structure] Why do patterns in moral judgements reflect legal principles humans are typically unaware of? (see *A Linguistic Analogy* in Lecture 02)
>
> [order-effects] Why are people's moral judgements about Switch and Drop subject to order-of-presentation effects (see *Framing Effects and Mikhail's Linguistic Analogy* in Lecture 02)[17]
>
> [dumbfounding-disengagement] Why are moral judgements sometimes, but not always, a consequence of reasoning from known principles? (see *Moral Dumbfounding* (section §2) and *Moral Disengagement: The Theory* (section §5))

8.2. Why The Four Puzzles Matter

To understand the roles of feeling and reasoning in moral intuitions and judgements, we must identify or create a theory that can solve the puzzles, is theoretically coherent and empirically motivated, and generates novel testable predictions.

---

[16] A linguistic analogy was introduced in *A Linguistic Analogy* in Lecture 02; we saw arguments against it in *Moral Disengagement: Significance* (section §7).

[17] Petrinovich & O'Neill (1996, Study 2); Wiegmann et al. (2012); Schwitzgebel & Cushman (2015)

Glossary

Cannibal 'Jennifer works in a medical school pathology lab as a research assistant. The lab prepares human cadavers that are used to teach medical students about anatomy. The cadavers come from people who had donated their body for the general use of the researchers in the lab. The bodies are normally cremated, however, severed cuts may be disposed of at the discretion of lab researchers, One night Jennifer is leaving the lab when she sees a body that is going to be discarded the next day. Jennifer was a vegetarian, for moral reasons. She thought it was wrong to kill animals for food. But then, when she saw a body about to be cremated, she thought it was irrational to waste perfectly edible meat. So she cut off a piece of flesh, and took it home and cooked it. The person had died recently of a heart attack, and she cooked the meat thoroughly, so there was no risk of disease.' (McHugh et al. 2023, supplementary materials; based on Haidt et al. 2000) 4–6

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.' 14

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.' 14

construct A factor postulated by a theory with the aim of explaining patterns of behaviour. Examples of constructs include moral conviction, moral disengagement and the moral foundations from Moral Foundations Theory. 11, 12

Drop A dilemma; also known as *Footbridge*. A runaway trolley is about to run over and kill five people. You can hit a switch that will release the bottom of a footbridge and one person will fall onto the track. The trolley will hit this person, slow down, and not hit the five people further down the track. Is it okay to hit the switch? 15

dual-process theory Any theory concerning abilities in a particular domain on which those abilities involve two or more processes which are dis-

tinct in this sense: the conditions which influence whether one mindreading process occurs differ from the conditions which influence whether another occurs. 13

Heinz 'In Europe, a woman was near death from a very bad disease, a special kind of cancer. There was one drug that the doctors thought might save her. It was a form of radium for which a druggist was charging ten times what the drug cost him to make. The sick woman's husband, Heinz, went to everyone he knew to borrow the money, but he could only get together about half of what it cost. He told the druggist that his wife was dying, and asked him to sell it cheaper or let him pay later. But the druggist said, "No, I discovered the drug and I'm going to make money from it." So, Heinz got desperate and broke into the man's store to steal the drug for his wife. The druggist had Heinz arrested and charged.' (McHugh et al. 2023, supplementary materials; based on Haidt et al. 2000) 4, 5

Incest 'Julie and Mark, who are brother and sister, are travelling together in France. They are both on summer vacation from college. One night they are staying alone in a cabin near the beach. They decide that it would be interesting and fun if they tried making love. At very least it would be a new experience for each of them. Julie was already taking birth control pills, but Mark uses a condom too, just to be safe. They both enjoy it, but they decide not to do it again. They keep that night as a special secret between them, which makes them feel even closer to each other' (McHugh et al. 2023, supplementary materials; based on Haidt et al. 2000) 4–6

moral disengagement Moral disengagement occurs when self-sanctions are disengaged from conduct. To illustrate, an executioner may avoid self-sanctioning for killing by reframing the role they play as 'babysitting' (Bandura 2002, p. 103). Bandura (2002, p. 111) identifies several mechanisms of moral disengagement: 'The disengagement may centre on redefining harmful conduct as honourable by moral justification, exonerating social comparison and sanitising language. It may focus on agency of action so that perpetrators can minimise their role in causing harm by diffusion and displacement of responsibility. It may involve minimising or distorting the harm that follows from detrimental actions; and the disengagement may include dehumanising and blaming the victims of the maltreatment.' 2, 7, 9, 11–14

moral dumbfounding 'the stubborn and puzzled maintenance of an [ethical] judgment without supporting reasons' (Haidt et al. 2000, p. 1). As

McHugh et al. (2017, p.) note, subsequent researchers have given different definitions of moral dumbfounding so that 'there is [currently] no single, agreed definition of moral dumbfounding.' I adopt the original authors' definition, as should you unless there are good reasons to depart from it. 2, 5, 6, 8, 14

moral intuition According to this lecturer, a person's intuitions are the claims they take to be true independently of whether those claims are justified inferentially. And a person's *moral* intuitions are simply those of their intuitions that concern ethical matters.

According to Sinnott-Armstrong et al. (2010, p. 256), moral intuitions are 'strong, stable, immediate moral beliefs.' 2, 7, 15

replicate To *replicate* a experiment is to attempt to repeat it with the aim of reproducing the original findings. Where the original findings are not found, it is called a *failed replication*.

A replication can be more or less *direct*; that is, it may adhere very closely to the original experiment, or it may include varations in the stimuli, subjects and settings. Very indirect replications are sometimes called *conceptual replications*. 5

Trolley A dilemma; also known as *Switch*. A runaway trolley is about to run over and kill five people. You can hit a switch that will divert the trolley onto a different set of tracks where it will kill only one. Is it okay to hit the switch? 4–6, 15

useful construct A *useful* construct is one that can explain an interesting range of target phenomena. 11

valid construct For the purposes of this course, a *valid* construct is one that can be measured using a tool (often a questionnaire) where there is sufficient evidence to conclude that the tool measures the construct. When used for cross-cultural comparisons, the tool should exhibit metric and scalar invariance (i.e. it should measure the same construct in the same way irrespective of which the culture participant belongs to).

Note that the term 'construct validity' is used in many different ways. It is probably best to try to understand it case-by-case—each time the term occurs, ask yourself what the researchers are claiming to have shown. If you do want an overview, Drost (2011) is one source. 11

References

Bandura, A. (2002). Selective Moral Disengagement in the Exercise of Moral Agency. *Journal of Moral Education*, *31*(2), 101–119.

Bandura, A., Barbaranelli, C., Caprara, G., & Pastorelli, C. (1996). Mechanisms of Moral Disengagement in the Exercise of Moral Agency. *Journal of Personality and Social Psychology*, *71*(2), 364–374.

Boardley, I. D. & Kavussanu, M. (2007). Development and Validation of the Moral Disengagement in Sport Scale. *Journal of Sport and Exercise Psychology*, *29*(5), 608–628.

Dahl, A. & Waltzer, T. (2018). Moral Disengagement As a Psychological Construct. *The American Journal of Psychology*, *131*(2), 240–246.

Drost, E. A. (2011). Validity and Reliability in Social Science Research. *Education Research and Perspectives*, *38*(1), 105–123.

Dwyer, S. (2009). Moral Dumbfounding and the Linguistic Analogy: Methodological Implications for the Study of Moral Judgment. *Mind & Language*, *24*(3), 274–296.

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.

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.

Guglielmo, S. (2018). Unfounded dumbfounding: How harm and purity undermine evidence for moral dumbfounding. *Cognition*, *170*, 334–337. [Online; accessed 2024-01-22].

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.

Haidt, J., Bjorklund, F., & Murphy, S. (2000). Moral dumbfounding: When intuition finds no reason.

Hindriks, F. (2014). Intuitions, Rationalizations, and Justification: A Defense of Sentimental Rationalism. *The Journal of Value Inquiry*, *48*(2), 195–216.

Hindriks, F. (2015). How Does Reasoning (Fail to) Contribute to Moral Judgment? Dumbfounding and Disengagement. *Ethical Theory and Moral Practice*, *18*(2), 237–250.

Hindriks, F. & Sauer, H. (2020). The mark of the moral: Beyond the sentimentalist turn. *Philosophical Psychology*, *33*(4), 569–591.

McAlister, A. L., Bandura, A., & Owen, S. V. (2006). Mechanisms of Moral Disengagement in Support of Military Force: The Impact of Sept. 11. *Journal of Social and Clinical Psychology*, *25*(2), 141–165.

McHugh, C., McGann, M., Igou, E. R., & Kinsella, E. L. (2017). Searching for Moral Dumbfounding: Identifying Measurable Indicators of Moral Dumbfounding. *Collabra: Psychology*, *3*(1), 23. [Online; accessed 2024-01-22].

McHugh, C., McGann, M., Igou, E. R., & Kinsella, E. L. (2020). Reasons or rationalizations: The role of principles in the moral dumbfounding paradigm. *Journal of Behavioral Decision Making*, *33*(3), 376–392. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/bdm.2167.

McHugh, C., McGann, M., Igou, E. R., & Kinsella, E. L. (2023). Cognitive Load Can Reduce Reason-Giving in a Moral Dumbfounding Task. *Collabra: Psychology*, *9*(1), 73818. [Online; accessed 2024-01-22].

McHugh, C., Zhang, R., Karnatak, T., Lamba, N., & Khokhlova, O. (2023). Just wrong? Or just WEIRD? Investigating the prevalence of moral dumbfounding in non-Western samples. *Memory Cognition*, *51*(5), 1043–1060. [Online; accessed 2024-01-22].

Mikhail, J. (2007). Universal moral grammar: Theory, evidence and the future. *Trends in Cognitive Sciences*, *11*(4), 143–152.

Nichols, S. (2004). *Sentimental rules: On the natural foundations of moral judgment*. Oxford: Oxford University Press.

Osofsky, M. J., Bandura, A., & Zimbardo, P. G. (2005). The Role of Moral Disengagement in the Execution Process. *Law and Human Behavior*, *29*(4), 371–393.

Pelton, J., Gound, M., Forehand, R., & Brody, G. (2004). The Moral Disengagement Scale: Extension with an American Minority Sample. *Journal of Psychopathology and Behavioral Assessment*, *26*(1), 31–39.

Petrinovich, L. & O'Neill, P. (1996). Influence of wording and framing effects on moral intuitions. *Ethology and Sociobiology*, *17*(3), 145–171.

Prinz, J. J. (2007). *The Emotional Construction of Morals*. Oxford: Oxford University Press.

Royzman, E. B., Kim, K., & Leeman, R. F. (2015). The curious tale of julie and mark: unraveling the moral dumbfounding effect. *Judgment & Decision Making*, *10*(4).

Royzman, E. B., Landy, J. F., & Goodwin, G. P. (2014). Are good reasoners more incest-friendly? Trait cognitive reflection predicts selective moralization in a sample of American adults. *Judgment and Decision Making*, *9*(3), 175.

Scanlon, T. M. (1998). *What we owe to each other*. Cambridge, Mass.: Belknap Press of Harvard University Press.

Schwitzgebel, E. & Cushman, F. (2015). Philosophers' biased judgments persist despite training, expertise and reflection. *Cognition*, *141*, 127–137.

Sinnott-Armstrong, W., Young, L., & Cushman, F. (2010). Moral intuitions. In J. M. Doris, M. P. R. Group, et al. (Eds.), *The moral psychology handbook* (pp. 246–272). Oxford: OUP.

Wiegmann, A., Okan, Y., & Nagel, J. (2012). Order effects in moral judgment. *Philosophical Psychology*, *25*(6), 813–836.

Wylie, D. (2021). The limits of moral dumbfounding. *Mind Language*, *36*(4), 610–626. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/mila.12391.