Lecture 02 : Moral Psychology

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

Tuesday, 19th January 2021

Contents

1. Why Is the Affect Heuristic Significant?

Why does it matter whether or not we use the Affect Heuristic? According to its defenders, it has implications for the foundations of ethics.

1.1. Two Implications of the Affect Heuristic for Ethics

First implication: 'if moral intuitions result from heuristics, [... philosophers] must stop claiming direct insight into moral properties' (Sinnott-Armstrong et al. 2010, p. 268).[1]

Second implication: 'Just as non-moral heuristics lack reliability in unusual situations, so do moral intuitions' (Sinnott-Armstrong et al. 2010, p. 268).

The second implication is relevant to evaluating objections to consequentialism:

> 'Critics often argue that consequentialism can't be accurate, because it implies moral judgments that are counter-intuitive, such as that we are morally permitted to punish an innocent person in the well-known example where this is necessary to stop riots and prevent deaths. With the heuristic model in hand, consequentialists can respond that the target attribute is having the best consequences, and any intuitions to the contrary result from substituting a heuristic attribute' (Sinnott-Armstrong et al. 2010, p. 269).

Wilson (who does not explicitly endorse the hypothesis that moral intuitions are a consequence of reliance on the Affect Hypothesis) makes an even stronger claim:

> 'ethical philosophers intuit the deontological canons of morality by consulting the emotive centers of their own hypothalamic-limbic system ... Only by interpreting the activity of the emotive centers as a biological adaptation can the meaning of the canons be deciphered' (Wilson 1975, p. 563 quoted in Haidt 2008, p. 68).

If this is right, you cannot understand ethics at all without knowledge of emotional processes. Wilson links this claim to a strong form of ethical pluralism:

[1] In *Moral Attributes Are Inaccessible* (section §4) we will see reason to doubt that this really is an implication. Note also that these philosophers' claim is quite narrow and does not bear directly on the view that ethical propositions may be self-evident if self-evidence is understood along the lines of Audi (2019). (Thanks to Paul Theo here.)

> 'a schedule of sex- and age-dependent ethics can impart higher genetic fitness than a single moral code which is applied uniformly to all sex-age groups. [...] no single set of moral standards can be applied to all human populations, let alone all sex-age classes within each population. To impose a uniform code is therefore to create complex, intractable moral dilemmas—these, of course, are the current condition of mankind' (Wilson 1975, pp. 563–4).

But should we accept any of these claims? Are they supported by evidence (or argument)?

2. Moral Intuitions and Emotions: Evaluating the Evidence

We have considered Schnall et al. (2008) as evidence for the idea that moral intuitions rely on the Affect Heuristic (as Sinnott-Armstrong et al (2010) propose). Whenever we encounter potential evidence, we should ask two questions of it. First, is it really evidence? Second, is it sufficient to justify us in accepting the claim we take it to be evidence for?

On this course you will be evaluating quite a lot of scientific evidence. As this is not something you are required to be familiar with doing before taking the course, I shall go through the process of evaluation quite slowly for the first time.

2.1. Step 0: Never Trust a Philosopher

This includes me, your lecturer. Always evaluate the evidence for yourself.

2.2. Step 1: Is It Really Evidence?

When faced with a potential piece of evidence, there are three questions you should always ask:

1. Has the study been replicated?
2. Are there similar studies? If so, are the findings convergent?
3. Has the study featured in a review? If so, does the review broadly support the findings of this study?

In the case of Schnall et al. (2008), I originally suggested (in the recording and the first version of these notes) that we answer as follows:

1. No, afaik this study has not been replicated. (This is fine; the only concern is when a study has been replicated and the replication failed.[2])
2. Yes, there are similar studies (e.g. Eskine et al. 2011); yes, these findings are convergent with those of Schnall et al. (2008).
3. Yes, the study has featured in at least one review (Chapman & Anderson 2013, p. 313). Yes, this review does broadly support the findings of Schnall et al. (2008).

However, since recording this section, I learned that Ugazio et al. (2012, Experiment 1a) report a failed partial replication of Schnall et al. (2008) (thank you Ollie!). Since these authors did not distinguish between high and low private body consciousness, the failure does not appear to be informative and does not undermine the main conclusion (whereas a further update, below, does undermine it).

The review mentioned in (3) provides strong support for the broad conclusion:

> 'To date, almost all of the studies that have manipulated disgust or cleanliness have reported effects on moral judgment. These findings strengthen the case for a causal relationship between disgust and moral judgment, by showing that experimentally evoked disgust—or cleanliness, its opposite—can influence moral cognition' (Chapman & Anderson 2013, p. 313).

At this point, it seems there is little doubt that we are right to take the findings of Schnall et al. (2008) as evidence. This is what I originally concluded, and what I say in the recording ('overwhelmingly yes' at 10:42). However, since then I realised that a meta-analysis by Landy & Goodwin (2015a) draws the opposite conclusion,[3] as does a recent study (Jylkkä et al. 2021; thank you Julina!). Authoritative commentaries by Giner-Sorolla et al. (2018, pp. 261–2) and Piazza et al. (2018) conclude that the available evidence is not strong.[4] If I were recording the lecture today, I would not be quite so bold. Overall we

[2] Some of the same authors pubilshed another study in the same year (Schnall et al. 2008) which an attempt to replicate has quite convincingly indicated that the effect is not powerful enough to have been discovered by the original study (Johnson et al. 2014). My recommendation is not to consider studies where there is an informative failure to replicate.

[3] Schnall et al. (2015) contest the latters' conclusions; Landy & Goodwin (2015b) make some interesting concessions in reply.

[4] McAuliffe (2019) also provides a review, but this is less nuanced. There are philosophical discussions, offering interestingly different perspectives, in May (2014), May (2018) and Kumar (2016). We will consider Kumar (2016) later in the context of dual process theories.

appear to have only weak evidence.

But there is a further question we should ask before accepting the Hypothesis about the Affect Heuristic.

2.3. Step 2: Is This Evidence Sufficient to Justify Accepting the Affect Heuristic?

Previously (in *Moral Intuitions and Emotions: Evidence* in Lecture 01) we considered supporting the Hypothesis that the Affect Heuristic is true by appeal to evidence for the correctness of one of its predictions. But this way of supporting the hypothesis has two weaknesses:

- it is post-hoc (the evidence for the prediction existed before the prediction was generated); and
- for all we know other predictions of the hypothesis may be falsified.

Neither weakness means that the evidence entirely fails to support the Hypothesis that the Affect Heuristic provides a correct account of moral intuitions. But these weaknesses do indicate that we require more robust support for the Hypothesis.

In searching for more robust support, we should consider the most successful arguments for heuristics (in reasoning generally, not in ethics specifically), and use these arguments as a model for what we would need to establish the Hypothesis about the Affect Heuristic.

2.4. Conclusions

Schnall et al. (2008) do provide good evidence.

We have sufficient evidence to conclude that feelings do influence moral intuitions (although this point will get further consideration in *PS: Does emotion influence moral judgment or merely motivate morally relevant action?* (section §7)).

But this evidence is not by itself sufficient to justify accepting the hypothesis that the Affect Heuristic provides a correct account of moral intuitions.

3. The Affect Heuristic and Risk: A Case Study

We need more evidence or argument to accept the the hypothesis that the Affect Heuristic explains moral intuitions. To identify what kind of evidence

or argument might work, we will consider a case from outside moral psychology whereas a similar hypothesis about a heuristic was successfully established.

By the end of this section you should better understand what kind of evidence might support the conclusion that people rely on a particular heuristic in making certain kinds of judgement.

In *Moral Intuitions and Emotions: Evaluating the Evidence* (section §2), we concluded that more evidence or argument would be needed to determine whether we should accept the hypothesis that the Affect Heuristic explains moral intuitions.

But what kind of evidence or argument could decide the issue? To answer this question, let us consider a case from outside moral psychology where a simililar hypothesis about a heuristic was successfully established.

3.1. A Case Study: Risk

Pachur et al. (2012) investigated how naive humans' unreflective judgements track three related attributes:

- frequency (which cause of death has a higher annual mortality rate?)
- risk (which cause of death represents a higher risk of dying from it?)
- Value of a Statistical Life, VSL (how much money should be spent to avoid one fatality due to this cause of death?)

You can see the actual frequencies (in Switzerland) and the subjects' median estimates of frequency for 24 types of cancer in Table 1 of Pachur et al. (2012).[5]

Since the attributes tracked were inaccessible to the subjects, they cannot

[5] Actual frequencies (in Switzerland) and subjects' median estimates of frequency for 24 types of cancer. *Source*:Pachur et al. (2012, Table 1).

	Annual mortality rate	*Mdn* frequency estimate	*Mdn* VSL	Affect (*M* dread score)	Availability (*M* number of instances in social network)
Penis cancer	10.2	40	10,000	4.65	0
Testicular cancer	17.2	109.5	25,000	4.55	0.03
Bone cancer	37.5	80	30,000	4.88	0.12
Thyroid cancer	69.3	80	30,000	4.44	0.03
Larynx cancer	94.2	50	20,000	3.92	0.24
Cancer of the connective tissue	94.3	45	17,500	4.4	0
Cancer of the gall bladder	196.5	30	11,010	4.31	0.09
Malignant melanoma (skin cancer)	242	50	17,102	3.59	0.24
Cervical cancer	295.8	95	20,000	4.39	0.15
Renal cancer	339.2	50	20,000	4.3	0.03
Cancer of the mouth and throat	351	60	12,609	3.91	0
Esophageal cancer	384.5	50	15,000	4.31	0.09
Rectal cancer	437.2	80	10,000	4.14	0
Bladder cancer	450.5	30	10,616	4.23	0.03
Ovarian cancer	453.2	100	20,000	4.43	0.03
Cancer of the nervous system	455	90	30,000	5.02	0.32
Hepatic cancer	513	110	10,000	4.04	0.18
Stomach cancer	572.2	115	20,000	4.22	0.18
Pancreatic cancer	897.8	60	17,501	4.43	0.26
Colon cancer	1,172.2	120	10,000	4.23	0.15
Prostate cancer	1,312.3	87	23,115	4.53	0.32
Leukemia and lymphoma	1,331.7	100	35,000	4.91	0.38
Breast cancer	1,347.3	200	40,000	4.24	0.79
Lung cancer	2,756	300	22,500	4	0.5

have been computing the attributues themselves. Instead they must have been computing something which, within limits, correlates with the attributes (like tracking toxicity by computing how smelling or tasting a potential food makes you feel; see *Two Questions about Moral Intuitions* in Lecture 01).

In this situation, there are at least two heuristics the subjects might use:

> *Availability Heuristic* The easier it is to bring a case of this cancer to mind, the more frequent or risky it is.
>
> *Affect Heuristic (for frequency and risk*[6] The more dread you feel when imagining it, the more frequent or risky it is.

Pachur et al. (2012) propose a hypothesis about how different attributes are tracked using different heuristics:

> Hypothesis: The Availability Heuristic dominates frequency judgements, whereas the Affect Heuristic dominates risk and VSL judgements.

This hypothesis generates a readily testable prediction:

> Prediction: Number of cases in a subject's social network will better predict frequency judgements, whereas feelings of dread will better predict risk and VSL judgements.

Pachur et al. (2012) tested these predictions. They found that:

> 'availability-by-recall offered a substantially better descriptive account than the affect heuristic when people judged deindividualized, statistical mortality rates. Affect, however, was at least on par with availability when people were asked to put a price tag on a single life saved from a risk, or when they were asked to indicate the perceived risk of dying' (p. 324).

I take these findings to provide a paradigm case where a hypothesis about a heuristic was successfully established. (But you should evaluate the evidence for yourself; see *Moral Intuitions and Emotions: Evaluating the Evidence* (section §2).)

How does this paradigm case compare with the argument and evidence we considered (see *Moral Intuitions and Emotions: Evidence* in Lecture 01) in the case of moral intuitions …

[6] Yes, it is potentially confusing that we are using the same term, 'Affect Heuristic' for a different heuristic. The common theme is tracking an attribute by computing how something makes you feel.

3.2. Applying the Case Study to Moral Intuitions

The argument of Pachur et al. (2012) involves two related features:

1. It is known in advance of gathering the evidence that the attribute tracked is inaccessible to the subjects (so could be not what they compute).
2. The evidence supports a view about *which* heuristic is used (Availablity or Affect), not *whether* a heuristic is used.

These two points indicate that we cannot take findings like those of Schnall et al. (2008) to establish conclusions about *whether* a heuristic is being used. Instead we need to establish, in advance of gathering such evidence, whether (or not) moral attributes are inaccesible.

Are moral attributes—like risk, frequency and VSL—inaccessible?

4. Moral Attributes Are Inaccessible

Sinnott-Armstrong et al. (2010, §2.1) argue that moral attributes are inaccessible. What is their argument, and does it work?

The canonical argument for the hypothesis that the Affect Heuristic explains moral intuitions depends on the premise that moral attributes are inaccessible (see *The Affect Heuristic and Risk: A Case Study* (section §**??**)).

Sinnott-Armstrong et al. (2010, p. 257) are blunt:

> 'Inaccessibility creates the need for a heuristic attribute'

These authors therefore claim to show (in Sinnott-Armstrong et al. 2010, §2.1) that:

> 'no plausible theory [of in what moral attributes consist] will make moral wrongness accessible' Sinnott-Armstrong et al. (2010, p. 257).

4.1. Aside on Implications of the Affect Heuristic

In *Why Is the Affect Heuristic Significant?* (section §1), I noted this claim:

> 'if moral intuitions result from heuristics, [... philosophers] must stop claiming direct insight into moral properties' (Sinnott-Armstrong et al. 2010, p. 268).

But we have seen (in *The Affect Heuristic and Risk: A Case Study* (section §3)) that the only available strategy for establishing for hypothesis that moral intuitions result from heuristics requires, *as a premise or lemma*, that moral

attributes are inaccessible. We cannot, therefore, regard this as an *implication* of that hypothesis. At least not unless we can find some further, as yet unknown, route to establishing it.

5. Moral Attributes Are Accessible

Here we consider arguments derived from Mikhail (2014) for the view that moral attributes are accessible.

Consider two questions of the same form but about different domains:

1. What do humans compute that enables them to track *moral* attributes?
2. What do humans compute that enables them to track *syntactic*[7] attributes?

A standard answer to the second question, (2), is: they compute the syntactic attributes themselves. Of course, humans are all, or mostly, unaware of computing syntactic attributes. But they do in fact do this, probably thanks to a language module.

Mikhail (2014) offers some considerations which can be used to argue for a parallel view about moral attributes:

> Humans track moral attributes by computing moral attributes.

This view would appear to imply that moral attributes are accessible.

5.1. Why Is This Relevant?

In *Moral Attributes Are Inaccessible* (section §4) we considered Sinnott-Armstrong et al. (2010, §2.1)'s argument that moral attributes are inaccessible.

I take that argument to depend on this premise:

> If it is hard to articulate some rules or how they apply to a situation, then any attribute characterised by those rules must be inaccessible.

Against this we may object:

[7] As an example of a syntactic attribute, consider *being a (grammatical) sentence.* For example, the sequence of words 'He is a waffling fatberg of lies' is a sentence whereas the sequence of words 'A waffling fatberg lies of he is' is not a sentence. These are syntactic attributes of the two sequences of words.

> It is hard to articulate syntactic rules, and to articulate how they apply to a sentence. But such rules characterise syntactic attributes, and syntactic attributes are not inaccessible. (Or if they are inaccessible, at least they are not inaccessible in any way that would support the argument for the hypothesis that the Affect Heuristic explains why humans have certain moral intuitions.[8])

This is not merely a hypothetical objection. For Mikhail (2007) argues that there is a relevant parallel between syntactic and ethical abilities.

5.2. What Is Mikhail's (Best) Argument?

1. 'adequately specifying the kinds of harm that humans intuitively grasp requires a technical legal vocabulary' (Mikhail 2007, p. 146)

   Therefore:

2. The abilities underpinning unreflective ethical judgements must involve analysis in accordance with rules.

   But:

3. Humans do not know the rules.

   Therefore:

4. The analysis is achieved by a modular process.

Mikhail's argument for the first premise that 'adequately specifying the kinds of harm that humans intuitively grasp requires a technical legal vocabulary' (Mikhail 2007, p. 146) depends on an analysis of pairs of dilemmas like the Trolley/Transplant pair presented in the recording. Many subjects make apparently inconsistent judgements when presented with such pairs of dilemmas; they appear to say that killing one to save five people is both permitted and impermissible. Mikhail argues that the inconsistency is merely apparent. For there is a morally significant difference between the dilemmas: one (Transplant) involves purposive battery while the other (Trolley) does not. This supports the idea that the pattern of judgements, far from being inconsistent, reflects the operation of principles and the identification of structure in the scenarios.[9]

---

[8] In Fodor (1983)'s characterisation of modularity, limited accessibility is one of the characteristics of modules. But note that limited accessibility is characteristic of the inner workings of a module, not of the judgements which modular processes influence.

[9] Mikhail (2014) provides more detail on the argument for this premise. (I also provide some detail in the recording.)

5.3. An Objection to Mikhail

Moral judgements are subject to order effects: which in a pair of dilemmas is presented first sometimes influences subjects' responses to the dilemmas (Petrinovich & O'Neill 1996, Study 2; Wiegmann et al. 2012). This is true even for professional philosophers (Schwitzgebel & Cushman 2015). No such effect is predicted by Mikhail's hypothesis that subjects' moral intuitions are a consequence of their correctly identifying structure and applying principles consistently.

Mikhail's hypothesis therefore at least requires qualification. This means his argument does not provide sufficient grounds to conclude that humans track moral attributes by computing moral attributes.

5.4. What Should We Conclude?

None of the arguments we have considered are sufficient to establish the view that moral intuitions are a consequence of a moral module.

But they are sufficient to show that the arguments considered in *Moral Attributes Are Inaccessible* (section §4) are not sufficient to establish that moral attributes are inaccessible.

This is an obstacle to establishing the hypothesis that the Affect Heuristic explains moral intuitions. For, as we saw (in *The Affect Heuristic and Risk: A Case Study* (section §3)) the best argument for that hypothesis depends on establishing that moral attributes are inaccessible.[10]

6. Conclusion: Two Puzzles

Our research on emotions and moral intuitions has left us with two puzzles. First, Why do feelings of disgust (and perhaps other emotions) moral intuitions? (And why do we feel disgust in response to moral transgressions?) Second, Why do patterns in moral intuitions reflect legal principles humans are typically unaware of?

6.1. Puzzle about Emotion

Why do feelings of disgust (and perhaps other emotions) influence moral intuitions? And why do we feel disgust in response to moral transgressions?

[10] This is not end of the story. We will find support for the view that, in some cases of moral intution, the moral attributes being tracked are inaccessible in *Why Is Moral Dumbfounding Significant?* in Lecture 03.

(This puzzle arises from *Moral Intuitions and Emotions: Evaluating the Evidence* (section §2).)

The second part of the puzzle is nicely articulated by Chapman & Anderson (2013, p. 317):

> 'What is the function of moral disgust? One of the most intriguing features of moral disgust is that it is not clear why it exists at all. Why should an emotion originating in defense against toxicity and disease be triggered by a social stimulus? The mystery deepens when we consider that human beings already have a social emotion that seems tailored to respond to moral wrongdoing, namely, anger [...]. Why then do we feel disgust in response to moral transgressions?'

6.2. Puzzle about Structure

Why do patterns in moral intuitions reflect legal principles humans are typically unaware of? (This puzzle arises from *Moral Attributes Are Accessible* (section §5).)

6.3. The Challenge We Face

We start from the question, What do adult humans compute that enables their moral intuitions to track moral attributes (such as wrongness)?

We have seen two candidate answers:

- they compute their emotional responses (Sinnott-Armstrong et al. 2010)
- they compute the moral attributes themselves (Mikhail 2007)

Each view is a response to a different puzzle grounded in an interesting, empirically-motivated theory. But neither seems fully able to explain all the puzzles.

Our task is to develop a theory that can solve the puzzles, is theoretically coherent and empirically motivated, and generates novel testable predictions.

7. PS: Does emotion influence moral judgment or merely motivate morally relevant action?

We have seen that manipulating emotion (in particular, disgust) can influence how people respond when asked to make moral judgement. Accord-

ing to Huebner et al. (2009), we should not conclude from this that emotion can influence moral judgement. They suggest that emotion may instead influence how scenarios are interpreted, how questions are understood; or it may 'act as a gain on what has already been conceived as a moral infraction (thereby, increasing the severity of the perceived wrong)' Are they right?

This question, which is a mini-essay task, is posed by Huebner et al. (2009). We saw (in *Moral Intuitions and Emotions: Evaluating the Evidence* (section §2)) that manipulating emotion (in particular, disgust) can influence how people respond when asked to make moral judgement.

This leads to a puzzle for advocates of views on which feelings play no role in moral intuitions (like that of Mikhail (2007); see *Moral Attributes Are Accessible* (section §5)). The puzzle is to explain why feelings of disgust (and perhaps other emotions) influence moral intuitions, and why we feel disgust in response to moral transgressions.

Can the puzzle be avoided? According to Huebner et al. (2009), findings' such as those of Schnall et al. (2008)

> 'fail to isolate the precise point at which emotion has a role in our moral psychology. ... emotional stimuli ... presented before the scenario is read could ... influence the interpretation of the scenario or the question. Or, emotion could act as a gain on what has already been conceived as a moral infraction (thereby, increasing the severity of the perceived wrong)' (Huebner et al. 2009, pp. 2–3).

Is this correct? If so, does it eliminate the need to respond to the puzzle about emotion?

More careful evaluations of the evidence are provided in reviews by Chapman & Anderson (2013) and Piazza et al. (2018), and a meta-analysis by Landy & Goodwin (2015a).[11] Interestingly, these reach quite different conclusions.

7.1. PS: What's Wrong with Huebner et al.'s Argument?

Anya challenged me to better explain why I conclude Huebner et al. (2009) fails to present a good challenge (thank you Anya!).

Suppose some researchers formulate a hypothesis H, generate some predictions and test them, and their predictions are all confirmed. No matter how often they do this, it will always be possible for a philosopher to identify an alternative hypothesis, H', which is consistent with all the observations

[11] Schnall et al. (2015) contest the latters' conclusions (see Landy & Goodwin (2015b) for a reply and Giner-Sorolla et al. (2018, pp. 261–2) for an evaluation).

made when testing the predictions. So we should not take the bare existence of an alternative hypothesis that is consistent with some observations to undermine the status of those observations as evidence for H.

This should be uncontroversial: it's barely saying more than that scientific arguments are not deductively valid.

My challenge to anyone who wants to use Huebner et al. (2009) is: Explain why your argument does more than establish that your opponents' observations are consistent with an alternative hypothesis, H', which is incompatible with your opponents' hypothesis.

Minimally, meeting this challenge requires showing that H and H' can be distinguished through readily testable predictions. If there's no prospect of us getting evidence to distinguish your opponents' hypothesis from your hypothesis, the distinction is unlikely to matter to us.

Ideally, meeting the challenge requires showing that your alternative hypothesis is theoretically coherent and empirically motivated.

To illustrate, Piazza et al. (2018) oppose the hypothesis that feelings of disgust influence moral judgements with the hypothesis that facts about what is disgusting play a conceptual role in categorising actions as good or bad.[12] But they do not offer this as a bare logical possibility. Instead they support their hypothesis with a mixture of argument and evidence. Their challenge should undermine our confidence in the hypothesis that feelings of disgust influence moral judgements to the extent that their alternative hypothesis is well supported.

8. Question Session 02

These are the recordings of the live online whole-class question session. They are usually available on the day after the session. (You may need to refresh this page to make them appear.)

[12] As they put it,

> 'the mind may use disgustingness to meaningfully separate and classify moral violations, and this organizing principle appears to operate separately from judgments of the acts' level of wrongdoing' (Piazza et al. 2018).

They distinguish this from a view on which it is feelings of disgust that influences judgement ('one may understand that an act is generally considered gross, and use that information to judge and act, regardless of whether one personally feels grossed out').

8.1. Direct Insight: Hannah's Question

Hannah asks, What 'direct insight' are other philosophers claiming to have into moral properties?

This question is about Sinnott-Armstrong et al. (2010, p. 268)'s rejection of direct insight.

Those authors mention Stratton-Lake (2002). I'm unsure exactly what they have in mind here as that's an edited collection with 12 chapters, but the introduction highlights something called *epistemological intuitionism*:

> 'Epistemological intuitionism is the view that certain moral propositions are self-evident—that is, can be know solely on the basis of an adequate understanding of them—and thus can be known directly by intuition' (Stratton-Lake 2002, p. 2).

As this book is not easily available (although you can ask the library to scan a chapter for you), I suggest considering Audi (2015), which contains the following claims:

> 'Intuition is a resource in all of philosophy, but perhaps nowhere more than in ethics' (p. 57).
>
> 'Episodic intuitions [...] can serve as data [...] ... beliefs that derive from them receive prima facie justification' (p. 65).
>
> 'self-evident propositions are truths meeting two conditions: (1) in virtue of adequately understanding them, one has justification for believing them [...]; and (2) believing them on the basis of adequately understanding them entails knowing them' (p. 65).

Paul Theo suggested a further way of thinking about 'direct insight' (thank you!):

> 'I think the other dominant way to justify direct insight (besides self-evidence views) would be the seeming state theory of intuition, which draws analogies between sense perception and moral intuition, and claims that both share a prima facie positive epistemic status.' See Pust (2019, §1.3).

Although I don't propose to consider such views in these lectures, you could reasonably consider them as part of your work for the course.

8.2. Defending Consequentialism: Jagoda's Second Question

Jagoda asks how accepting the hypothesis that the Affect Heuristic explains moral intuitions could be part of a defence of consequentialism. (This relates

to *Why Is the Affect Heuristic Significant?* (section §1).)

Baldouin offers a quote that is relevant to this question:

> 'Critics often argue that consequentialism can't be accurate, because it implies moral judgments that are counter-intuitive, such as that we are morally permitted to punish an innocent person in the well-known example where this is necessary to stop riots and prevent deaths. With the heuristic model in hand, consequentialists can respond that the target attribute is having the best consequences, and any intuitions to the contrary result from substituting a heuristic attribute' (Sinnott-Armstrong et al. 2010, p. 269).

As Baldouin notes, the suggestion is that the hypothesis about the Affect Heuristic can play a role in responding to some objections to consequentialism. As far as I know, neither Sinnott-Armstrong et al. (2010) nor others are suggesting that there is a role for this hypothesis in establishing, positively, that consequentialism is true.

8.3. Acceptance: Jagoda's 11:55 question

Jagoda opens by asking, 'Is acceptance a moral attribute?'.

No. Wrongness is a moral attribute; perhaps also when we talk of actions being harmful, unfair, disloyal, disrespectful or impure we are attributing moral attributes. Maybe there are other moral attributes. But acceptance seems like the wrong kind of thing to be a moral attribute.

Jagoda's question is related to *Moral Attributes Are Inaccessible* (section §4), and in particular to the fourth candidate for moral wrongness. According to this candidate, for an act to be morally wrong is for it to violate a rule that all impartial, rational people would accept. Jagoda objects:

> Through accepting something we are making the decision that we either agree with it or have no objections to it and we surely do this through making some sort of judgement about the thing we are accepting and why we accept it. Therefore, [the fourth candidate would imply that] we cannot have people who are both impartial and yet accept something (a rule)?'

I think Jagoda is right that these rough-and-ready charicatures of how to characterise moral wrongness would need substantial refinement to be plausible. They are just props in an argument for the claim that moral attributes are inaccessible.

I argued that this argument fails (in *Moral Attributes Are Accessible* (section §5)[13]). Perhaps Jagoda's objection adds support to that view.

8.4. Metaethics and Epistemology: Svenja's and Adam's questions

Svenja asks,

> When we talk about an action having the 'attribute of moral wrongness' and this attribute being accessible to us or not, are we assuming that such an attribute exists independently of us and that whether or not an action has this attribute is an objective fact?

In reply I offer a comparison between *wrongness* and *blueness*.[14] In both cases, I'm unsure whether we are assuming that the attribute exists independently of us, nor whether there are objective facts about it. But if there is a problem here, I think it's unlikely to be a specifically ethical problem.

Adam follows up by objecting,

> Do Sinnott-Armstrong et al. (2010) believe moral attributes are inaccessible because they are inherently impossible to gain direct insight into or just difficult? If they are inherently impossible, we have never gained direct insight, so on what basis can we say that we have any grip on moral attributes whatsoever? If they are difficult to compute, how difficult? It looks to me like Sinnott-Armstrong et al. (2010) is saying that gaining insight into moral attributes is in principle possible but in practice impossible, in which case, the problem remains.

What I should have done in response to this part of Adam's objection is referred back to the comparison between moral attributes and risk (in *The Affect Heuristic and Risk: A Case Study* (section §3)). Risks are often inaccessible to non-experts who have not suffered personal tragedies. But such people can become experts, of course. Similarly, I think the only premise Sinnott-Armstrong et al. (2010) need is that the moral attributes relevant to particular

[13] Note that this section does not provide sufficient reason to conclude that moral attributes *are* accessible. It's concluion is that the arguments considered to not provide sufficient reason for the view that moral are inaccessible nor for the view that moral attributes are accessible.

[14] I cited Roberson & Hanley (2010) as they provide background on Berinmo colour words, and Witzel & Gegenfurtner (2016) as an example of evidence that there is a link between which colour words you have and which categorical colour properties you can discriminate. Witzel & Gegenfurtner (2018) offer a useful review.

actions are inaccessible to most participants in experiments involving those actions.

Adam objects, further, that

> Sinnott-Armstrong et al. (2010) seem to be saying that the difficulty of knowing objective moral attributes is just a problem for the reliability of moral intuitions. However, if no one currently alive can gain insights about the moral properties of an event doesn't that raise much more fundamental questions about what right [justification?] we have to even assign moral properties to events in the first place?

I think this is wrong. They are not saying that moral intuitions are unreliable in general. Given some background assumptions about the limits of reason, their view is compatible with the claim that in many familiar situations, moral intuitions are a highly reliable basis for reaching a conclusion (and may be the most reliable basis available). Their hypothesis about the Affect Heuristic only implies that moral intuitions are likely to be unreliable in unfamiliar situations.

So I agree with Adam that we are skirting questions about justification for propositions about moral attributes. But I take the opposite view: something like the hypothis that moral intutions are a consequence of the Affect Hypothesis may will provide the basis for a plausible account of how humans can know that some actions are right and others wrong.

Glossary

Affect Heuristic In the context of moral psychology, the Affect Heuristic is this principle: 'if thinking about an act [...] makes you feel bad [...], then judge that it is morally wrong' (Sinnott-Armstrong et al. 2010). These authors hypothesise that the Affect Heuristic explains moral intuitions.

A different (but related) Affect Heurstic has also be postulated to explain how people make judgements about risky things are: The more dread you feel when imagining an event, the more risky you should judge it is (see Pachur et al. 2012, which is discussed in 3). 3, 4, 6–9, 12, 16, 19

heuristic A *heuristic* links an inaccessible attribute to an accessible attribute such that, within a limited but useful range of situations, someone could track the inaccessible attribute by computing the accessible attribute. 8

inaccessible An attribute is *inaccessible* in a context just if it is difficult or impossible, in that context, to discern substantive truths about that attribute. For example, in ordinary life and for most people the attribute *being further from Kilmery (in Wales) than Steve's brother Matt is* would be inaccessible.

See Kahneman & Frederick (2005, p. 271): 'We adopt the term accessibility to refer to the ease (or effort) with which particular mental contents come to mind.' 7, 9, 10, 12, 17, 18

module A *module* is standardly characterised as a cognitive system which exhibits, to a significant degree, a set of features including domain specificity, limited accessibility, and information encapsulation. Contemporary interest in modularity stems from Fodor (1983). Note that there are now a wide range of incompatible views on what modules are and little agreement among researchers on what modules are or even which features are characteristic of them. 10, 12

moral intuition According to this lecturer, moral intuitions are unreflective ethical judgements.

According to Sinnott-Armstrong et al. (2010, p. 256), moral intuitions are 'strong, stable, immediate moral beliefs.' 4, 6–8, 12–14, 16, 19

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*. 4

track For a process to *track* an attribute is for the presence or absence of the attribute to make a difference to how the process unfolds, where this is not an accident. (And for a system or device to track an attribute is for some process in that system or device to track it.)

Tracking an attribute is contrasted with *computing* it. Unlike tracking, computing typically requires that the attribute be represented. (The distinction between tracking and computing is a topic of **??**.) 7, 10

Transplant A dilemma. Five people are going to die but you can save them all by cutting up one healthy person and distributing her organs. Is it ok to cut her up? 11

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? 11

unfamiliar problem An unfamiliar problem (or situation) is one 'with which we have inadequate evolutionary, cultural, or personal experience' (Greene 2014, p. 714). 3, 19

References

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

Audi, R. (2019). Understanding, Self-Evidence, and Justification. *Philosophy and Phenomenological Research*, *99*(2), 358–381.

Chapman, H. A. & Anderson, A. K. (2013). Things rank and gross in nature: A review and synthesis of moral disgust. *Psychological Bulletin*, *139*(2), 300–327.

Eskine, K. J., Kacinik, N. A., & Prinz, J. J. (2011). A Bad Taste in the Mouth: Gustatory Disgust Influences Moral Judgment. *Psychological Science*, *22*(3), 295–299.

Fodor, J. (1983). *The Modularity of Mind: an Essay on Faculty Psychology*. Bradford book. Cambridge, Mass ; London: MIT Press.

Giner-Sorolla, R., Kupfer, T., & Sabo, J. (2018). What Makes Moral Disgust Special? An Integrative Functional Review. In J. M. Olson (Ed.), *Advances in Experimental Social Psychology*, volume 57 (pp. 223–289). Academic Press.

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

Haidt, J. (2008). Morality. *Perspectives on Psychological Science*, *3*(1), 65–72.

Huebner, B., Dwyer, S., & Hauser, M. (2009). The role of emotion in moral psychology. *Trends in Cognitive Sciences*, *13*(1), 1–6.

Johnson, D. J., Cheung, F., & Donnellan, M. B. (2014). Does Cleanliness Influence Moral Judgments? *Social Psychology*, *45*(3), 209–215.

Jylkkä, J., Härkönen, J., & Hyönä, J. (2021). Incidental disgust does not cause moral condemnation of neutral actions. *Cognition and Emotion*, *35*(1), 96–109.

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

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

Landy, J. F. & Goodwin, G. P. (2015a). Does incidental disgust amplify moral judgment? A meta-analytic review of experimental evidence. *Perspectives on Psychological Science*, *10*(4), 518–536.

Landy, J. F. & Goodwin, G. P. (2015b). Our conclusions were tentative, but appropriate: A reply to schnall et al.(2015). *Perspectives on Psychological Science*, *10*(4), 539–540.

May, J. (2014). Does Disgust Influence Moral Judgment? *Australasian Journal of Philosophy*, *92*(1), 125–141.

May, J. (2018). *Regard for Reason in the Moral Mind*. Oxford University Press.

McAuliffe, W. H. B. (2019). Do emotions play an essential role in moral judgments? *Thinking & Reasoning*, *25*(2), 207–230.

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

Mikhail, J. (2014). Any Animal Whatever? Harmful Battery and Its Elements as Building Blocks of Moral Cognition. *Ethics*, *124*(4), 750–786.

Pachur, T., Hertwig, R., & Steinmann, F. (2012). How Do People Judge Risks: Availability Heuristic, Affect Heuristic, or Both? *Journal of Experimental Psychology: Applied*, *18*(3), 314–330.

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

Piazza, J., Landy, J. F., Chakroff, A., Young, L., & Wasserman, E. (2018). What disgust does and does not do for moral cognition. In N. Strohminger & V. Kumar (Eds.), *The Moral Psychology of Disgust* (pp. 53–81). Rowman & Littlefield International.

Pust, J. (2019). Intuition. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2019 ed.). Metaphysics Research Lab, Stanford University.

Roberson, D. & Hanley, R. (2010). Relatively speaking. *Words and the Mind*, *1*(9), 183–199.

Schnall, S., Benton, J., & Harvey, S. (2008). With a Clean Conscience: Cleanliness Reduces the Severity of Moral Judgments. *Psychological Science*, *19*(12), 1219–1222.

Schnall, S., Haidt, J., Clore, G. L., & Jordan, A. H. (2008). Disgust as Embodied Moral Judgment. *Personality and Social Psychology Bulletin*, *34*(8), 1096–1109.

Schnall, S., Haidt, J., Clore, G. L., & Jordan, A. H. (2015). Landy and Goodwin (2015) Confirmed Most of Our Findings Then Drew the Wrong Conclusions. *Perspectives on Psychological Science*, *10*(4), 537–538.

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.

Stratton-Lake, P. (Ed.). (2002). *Ethical Intuitionism: Re-Evaluations*. Oxford: Clarendon Press.

Ugazio, G., Lamm, C., & Singer, T. (2012). The Role of Emotions for Moral Judgments Depends on the Type of Emotion and Moral Scenario. *Emotion*, *12*(3), 579–590.

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

Wilson, E. O. (1975). *Sociobiology: The New Synthesis*. Cambridge, Mass: Belknap Press of Harvard University Press.

Witzel, C. & Gegenfurtner, K. R. (2016). Categorical perception for red and brown. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(4), 540–570.

Witzel, C. & Gegenfurtner, K. R. (2018). Color Perception: Objects, Constancy, and Categories. *Annual Review of Vision Science*, *4*(1), 475–499.