Lecture 02 : Moral Psychology

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

Thursday, 16th January 2025

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]   I am doubtful that this really is an implication: if I understand Sinnott-Armstrong et al. (2010)'s position, there would be no need for heuristics at all if moral attributes were not inaccessible—so it appears to me as if the inaccessibility of moral attributes is a premise rather than 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)?

1.2.   Background: Understanding Heuristics

To gain a better understanding of heuristics, it may be helpful to consider a nonmoral case where we have good evidence that heuristics matter.

Pachur et al. (2012) investigated how naive humans' answer track three questions:

- frequency—Which cause of death has a higher annual mortality rate?
- risk—Which cause of death represents a higher risk of dying from it?[2]
- Value of a Statistical Life (VSL)—How much money should be spent to avoid one fatality due to this cause of death?[3]

You can see the actual frequencies (in Switzerland) and the subjects' median estimates of frequency for various causes of death in Table 2 of Pachur et al. (2012).

What did the subjects compute that enabled them to answer questions about frequency, risk and the value of a statistical life? Since these attributes tracked were inaccessible to the subjects, they cannot 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 *Moral Intuitions and an Affect Heuristic* 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.

---

[2]   You might reasonably hold that frequency and risk are not distinct. But you will see that people tend to make different judgements in response to the risk and frequency questions.

[3]   The lecture omits discussion of VSL for simplicity.

> *Affect Heuristic (for frequency and risk*[4]*)* 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 deindivid-
> ualized, 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).

These findings provide a paradigm case where a hypothesis about a heuristic was successfully established.

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.

---

[4]   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.

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

Before we rely on the findings of a study, we should ideally have positive answers to these questions. (Perhaps we do not need a successful replication, but if so there should at least not be unexplained failures to replicate the study.)

In the case of Schnall et al. (2008), the answers to these questions are quite complicated:

1. 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).[5] But, as you can see below, there are also reviews which support a view incompatible with these findings.

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. Experiment 1 of Schnall et al. (2008) is actually a successful conceptual replication of Wheatley & Haidt (2005).

    Ugazio et al. (2012) report an unsuccessful attempt to replicate Experiment 1 of Schnall et al. (2008). Although this failed replication may weaken our confidence in Schnall

---

[5]   Chapman & Anderson (2013, p. 313) 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.' Note, however, that in later work Chapman (2018, pp. 73–4) revises this view in light of new evidence: 'a recent meta-analysis of disgust induction studies suggests that incidental disgust has at best a small effect on moral judgment (Landy & Goodwin 2015a).'

et al. (2008)'s findings, note that, as Ugazio et al. (2012, p. 589) report, there is an extraneous weakness in attempted replication which explain may the failure to replicate.[6]

Johnson et al. (2016) report an convincing failure to replicate Experiment 3 (the one where disgust is induced by having participants recall a disgusting event in their lives.)[7]

At this point, it seems we can take the findings of Schnall et al. (2008) as evidence.[8] However, a more recent meta-analysis by Landy & Goodwin (2015a) draws the opposite conclusion,[9] 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.[10] Indeed, Piazza et al. (2018, p. 54) argue that 'robust evidence is lacking for a unique effect of disgust on moral judgment.'

This view might be further supported by Yang et al. (2013)'s findings which appear to show that moral evaluations occur before disgust responses to a scenario. They suggest that 'the temporal priority of morality over physical disgust further suggests that the physical disgust emotion may not be necessary for people to make moral judgments' (p. 5).

What should we conclude? Without closer evaluation of more experimental findings (which is surely worthwhile, although not for everybody), we

---

[6] Some of the same authors pubilshed another study in the same year (Schnall et al. 2008). Johnson et al. (2014) attempted to replicate this other study. Those authors' results convincingly indicate that the effect is not powerful enough to have been discovered by the original study. This is an informative failure to replicate. My recommendation is therefore not to consider (Schnall et al. 2008) as evidence.

[7] Johnson et al. (2016)'s primary focus is where individual variation in disgust sensitivity (as measured by private body consciousness) mediates the influence of disgust on moral judgement. They conclude that 'written disgust manipulations did not impact moral judgments, nor was this effect moderated by individual differences in sensitivity to internal bodily sensations' (Johnson et al. 2016, p. 6).

[8] This is what I originally concluded, and what I say in the recording ('overwhelmingly yes' at 10:42). When I eventually update the recording, I will not say that.

[9] Schnall et al. (2015) contest the latters' conclusions; Landy & Goodwin (2015b) make some interesting concessions in reply. See also Giner-Sorolla et al. (2018, pp. 261–2) for an evaluation. Interestingly, a more recent meta-analysis by Salvo et al. (2025). reaches quite different conclusions. While I am not certain why their findings are so different from Landy & Goodwin (2015a)'s, at least part of the explanation appears to be that Salvo et al are not focussed on incidental disgust and not as rigourous in attempting to consider the effects of confounds. (Insights on this are welcome.)

[10] 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.

should be cautious in taking Schnall et al. (2008) or similar studies as providing strong evidence that experimentally induced extraneous disgust makes people harsher in their moral judgements.

We should not infer that emotions and feelings do not influence moral intuitions at all. There are other ideas about how disgust and other feelings could influence emotion which are supported by other sources of evidence (for instance, Piazza et al. 2018).

2.3.   Conclusion

How, if at all, do emotion or feelings influence moral intuitions? Since the Affect Heuristic provides a direct and bold answer to this question, we have been concerned to identify and evaluate evidence for Affect Heuristic. One key piece of evidence cited by Sinnott-Armstrong et al. (2010) is Schnall et al. (2008). Although initially convincing, considering a wider range of research indicates that this is probably not strong evidence for the Affect Heuristic.

This may motivate us to consider other possible sources of evidence for the Affect Heuristic. Or it may motivate us to consider other theories about how emotions or feelings influence moral intuitions.[11]

3.   A Linguistic Analogy

What do humans compute that enables them to track moral attributes? In this section we introduce a second hypothesis which answers this question, one based on an analogy between ethical and linguistic abilities. The hypothesis is due to Mikhail (2014). Considering the hypothesis also provides an argument for the view that moral attributes are accessible.

Several researchers have developed theories about humans' ethical abilities based on analogies with their linguistic abilities (including Mikhail 2007, Dwyer 2009 and Roedder & Harman 2010).

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

> 1.   What do humans compute that enables them to track *moral* attributes?

---

[11]   Possible sources include Ugazio et al. (2012). They aim to 'uncover the mechanisms by which emotions exert their influence on moral judgments' (p. 587) by comparing the effects of different emotions—anger and disgust—on responses to four scenarios involving moral violations. Further, Decety & Cacioppo (2012) argue that 'moral reasoning involves a complex integration between emotion and cognition that gradually changes with age.'

2. What do humans compute that enables them to track *syntactic*[12] 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 in much the way that they track linguistic attributes (which perhaps involves a language module).

This is an alternative to Sinnott-Armstrong et al. (2010, §2.1)'s hypothesis about the Affect Heuristic.

3.1.   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 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

---

[12] 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.

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.[13]

3.2. 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.

3.3. What Should We Conclude?

None of the arguments we have yet considered are sufficient to establish the view that moral intuitions are a consequence of a moral module. So while the idea that there is an analogy between ethical and linguistic abilities remains intriguing, we are not in a position to accept or reject it without further arguments or discoveries.

3.4. Appendix: What Are Modules?

They are 'the psychological systems whose operations present the world to thought'; they 'constitute a natural kind'; and there is 'a cluster of properties that they have in common' (Fodor 1983, p. 101):

- domain specificity (modules deal with 'eccentric' bodies of knowledge)
- limited accessibility (representations in modules are not usually inferentially integrated with knowledge)
- informational encapsulation (modules are unaffected by general knowledge or representations in other modules)

---

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

- innateness (roughly, the information and operations of a module not straightforwardly consequences of learning; but see Samuels (2004)).

4.   Framing Effects and Mikhail's Linguistic Analogy (optional)

Mikhail (2007) argues for an analogy between ethical and linguistic abilities on the grounds that patterns in humans' moral intutions reflect legal principles they are unaware of. One challenge to this argument arises from evidence that moral intuitions are subject to framing effects.

This section is not included in the lecture in 2024–5. If you are a working on Mikhail's argument (see A Linguistic Analogy *(section §unresolved xref to unit:moral_attributes_accessible_pro)), the notes may help you. We will cover framing effects in Part III of the course.*

4.1.   What Is a Framing Effect?

Suppose you are asked to judge whether an object is near or far from you. You might be surprised to discover that your judgements can be influenced by whether another person is in the scene and able to interact with the object (Fini et al. 2015). After all, the judgement you are making is supposed to be about the distance between you and an object; the distance from another person and that person's ability to interact with the object are irrelevant considerations.

This an example of a framing effect: task-irrelevant features of a situation systematically influence your performance.

4.2.   Are Philosophers Subject To Framing Effects When Considering Ethical Scenarios?

Schwitzgebel & Cushman (2015) show that philosophers are subject to order-of-presentation effects (they make different judgements depending on which order trolley scenarios are presented).

Wiegmann et al. (2020) show that philosophers are subject to irrelevant additional options: like lay people, philosophers will more readily endorsing killing one person to save nine when given five alternatives than when given six alternatives. (These authors also demonstrate order-of-presentation effects.)

Wiegmann & Horvath (2021) show that they philosophers are subject to the 'Asian disease' framing used in a famous earlier study (Tversky & Kahneman 1981). (They also find an indication that philosophers, although susceptible to other framing effects, may be less susceptible than lay people to four other framing effects, including whether an outcome is presented as a loss or a gain (which they term 'Focus').)

4.3.   What do Framing Effects Show?

According to Kahneman (2013), there are *some instances*[14] in which

> 'there is no underlying preference that is masked or distorted by the frame. Our preferences are about framed problems, and our moral intuitions are about descriptions, not about substance' (Kahneman 2013).

From the existence of framing effects, Rini (2013, p. 265) also draws a strong conclusion:

> 'Our moral judgments are apparently sensitive to idiosyncratic factors, which cannot plausibly appear as the basis of an interpersonal normative standard. [...] we are not in a position to introspectively isolate and abstract away from these factors. Worse yet, even when we think that we have achieved normative abstraction, we may only erroneously conclude that we have succeeded.'

Should we accept these strong conclusions? Perhaps there are good arguments for them, but we cannot draw either Kahneman's or Rini's conclusion directly from the mere existence of framing effects. Consider order-of-presentation effects. Wiegmann & Waldmann (2014) offer evidence for the theory that this effect is a consequence of one scenario selectively highlighting an aspect of the causal structure of another scenario. If this is correct, we might think that the order-of-presentation effect does not show that moral intuitions are not about substance after all.

4.4.   Conclusion

Mikhail (2007) offered an argument for his Linguistic Analogy based the claim that there is a pattern in humans' moral intuitions: they relfect legal principles such as a ban on purposive battery.

---

[14] Kahneman (2013) is making this claim for Schelling's child exemptions in the tax code example and also the 'Asian disease' framing effect.

However we have seen that these moral intuitions are subject to framing effects.

We should therefore not accept Mikhail (2007)'s argument unless there is good reason to suppose that the pattern Mikhail identifies (see table 2 in Mikhail 2007) are not distorted by framing effects.

5.   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) influence 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?

5.1.   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? (This puzzle arises from *Moral Intuitions and Emotions: Evaluating the Evidence* (section §**unresolved xref to unit:moral_intuitions_evidence_evaluation**).)

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?'

5.2.   Puzzle about Structure

Why do patterns in moral intuitions reflect legal principles humans are typically unaware of? (This puzzle arises from *A Linguistic Analogy* (section §**unresolved xref to unit:moral_attributes_accessible_pro**).)

5.3.   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.

6.   Question Session 02

There are no question sessions this year, but some of the notes from previous years are still relevant. These are included here.

6.1.   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 justifica-
> tion 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 in-
> tuition, which draws analogies between sense perception and
> moral intuition, and claims that both share a prima facie posi-
> tive 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.

6.2.   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 re-
lates to *Why Is the Affect Heuristic Significant?* (section §**unresolved xref to
unit:affect_heuristic_significance**).)

Baldouin offers a quote that is relevant to this question:

> 'Critics often argue that consequentialism can't be accurate, be-
> cause 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 ri-
> ots and prevent deaths. With the heuristic model in hand, con-
> sequentialists 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 consequential-
ism. 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.

7. 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.

This section is not part of the spoken lecture but may be useful as background.

In *Moral Intuitions and Emotions: Evaluating the Evidence* (section §**unresolved xref to unit:moral_intuitions_evidence_evaluation**), 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.

7.1. A Case Study: Risk

Pachur et al. (2012) investigated how naive humans' answer track three questions:

- frequency—Which cause of death has a higher annual mortality rate?
- risk—Which cause of death represents a higher risk of dying from it?[15]
- 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).[16]

Since the attributes tracked were inaccessible to the subjects, they cannot have been computing the attributues themselves. Instead they must have

---

[15] You might reasonably hold that frequency and risk are not distinct. But you will see that people tend to make different judgements in response to the risk and frequency questions.

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

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 *Moral Intuitions and an Affect Heuristic* 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[17])* 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).

	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

1).

[17] 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.

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 §**unresolved xref to unit:moral_intuitions_evidence_evaluation**).)

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 ...

7.2.   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?

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. 3, 5, 8, 9, 15–17

Asian disease   A disease will kill 600 people for sure without an intervention. You are a decision maker tasked with choosing between two interventions. Your choice can be framed in two ways. Frame 1: Either save 200 people for sure, or else take a one in three chance that everyone will be saved with a two in three chance that no one will be saved. Frame 2: Either allow 400 people to die for sure, or else take a one

in three chance that nobody will die and a two in three chance that everyone will die. (Tversky & Kahneman 1981) 12

domain specific A process is domain specific to the extent that there are limits on the range of functions its outputs typically serve. Domain-specific processes are commonly contrasted with general-purpose processes. 10, 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. 4, 17

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.' 4, 8, 16, 18, 19

informational encapsulation One process is informationally encapsulated from some other processes to the extent that there are limits on the one process' ability to consume information available to the other processes. (See Fodor 1983; Clarke 2020, pp. 5ff.) 10, 19

innate Not learned. While everyone disagrees about what innateness is (see Samuels 2004), on this course a cognitive ability is innate just if its developmental emergence is not a direct consequence of data-driven learning. 11

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 informational 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. 9, 10

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.' 5, 10, 13, 15, 16, 18

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*. 6

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 **??**.) 4, 8, 9, 16

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? 9, 10

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? 9, 10

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

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. (2018). *A component process model of disgust, anger, and moral judgment*, volume 70, (pp. 70—80). New York: Guilford Publications.

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.

Clarke, S. (2020). Cognitive penetration and informational encapsulation: Have we been failing the module? *Philosophical Studies*.

Decety, J. & Cacioppo, S. (2012). The speed of morality: A high-density electrical neuroimaging study. *Journal of Neurophysiology*, *108*(11), 3068–3072.

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

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.

Fini, C., Brass, M., & Committeri, G. (2015). Social scaling of extrapersonal space: Target objects are judged as closer when the reference frame is a human agent with available movement potentialities. *Cognition*, *134*, 50–56.

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.

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

Johnson, D. J., Wortman, J., Cheung, F., Hein, M., Lucas, R. E., Donnellan, M. B., Ebersole, C. R., & Narr, R. K. (2016). The effects of disgust on moral judgments: Testing moderators. *Social Psychological and Personality Science*, *7*(7), 640–647.

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

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

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.

Rini, R. A. (2013). Making psychology normatively significant. *The Journal of ethics*, *17*(3), 257–274.

Roedder, E. & Harman, G. (2010). Linguistics and moral theory. In J. M. Doris, M. P. R. Group, et al. (Eds.), *The moral psychology handbook* (pp. 273–296). Oxford: Oxford University Press.

Salvo, G., Ottaviani, C., & Mancini, F. (2025). Bidirectional interplay of disgust and morality: Meta-analytic investigations. *Personality and Individual Differences*, *236*, 113032. [Online; accessed 2025-02-20].

Samuels, R. (2004). Innateness in cognitive science. *Trends in Cognitive Sciences*, *8*(3), 136–41.

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: Oxford University Press.

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

Tversky, A. & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453–458.

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.

Wheatley, T. & Haidt, J. (2005). Hypnotic Disgust Makes Moral Judgments More Severe:. *Psychological Science*, *16*(10), 780–784.

Wiegmann, A. & Horvath, J. (2021). Intuitive Expertise in Moral Judgements. *Australasian Journal of Philosophy*, *100*(2), 342–359.

Wiegmann, A., Horvath, J., & Meyer, K. (2020). Intuitive expertise and irrelevant options. *Oxford studies in experimental philosophy*, *3*, 275–310.

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

Wiegmann, A. & Waldmann, M. R. (2014). Transfer effects between moral dilemmas: A causal model theory. *Cognition*, *131*(1), 28–43.

Wilson, E. O. (1975). *Sociobiology: The New Synthesis*. Cambridge, Massachusetts: Belknap Press.

Yang, Q., Yan, L., Luo, J., Li, A., Zhang, Y., Tian, X., & Zhang, D. (2013). Temporal Dynamics of Disgust and Morality: An Event-Related Potential Study. *PLOS ONE*, *8*(5), e65094. [Online; accessed 2025-03-04].