Phil 134/ 234 G
Moral Psych: S 21

Prof. Aaron Z. Zimmerman
May 4, 2021

Handout #6: The Cognitive Neuroscience of Moral Judgment

1. The Relationship between Cognitive Neuroscience and Moral/Political Theory

"The cognitive science of morality (CSM)… is not only uncovering the psychological structures that underlie moral judgment but, increasingly, also their neural underpinning—utilizing, in this connection, advances in functional neuroimaging, brain lesion studies, psychopharmacology, and even direct stimulation of the brain. Evidence from such research has been used not only to develop grand theories about moral psychology, but also to support ambitious normative arguments."

The Relevance Thesis: CSM is "highly relevant" to moral/political theory: "Some assert that empirical evidence could resolve longstanding ethical debates" (see e.g. Churchland, 2011; Greene, 2008, 2016).

The Irrelevance Thesis: CSM is not relevant to moral/political theory. "Cognitive neuroscience has no normative significance whatsoever" (Berker, 2009).

The Probably Mostly Irrelevant Thesis endorsed by J. Demaree-Cotton and G. Kahane (D&K): "Focusing on the issue of the reliability of our moral judgments, we shall suggest that neuroscientific findings have limited epistemic significance considered on their own; they are likely to make an epistemic difference only when "translated" into higher-level psychological claims."

2. Is Neuroscience Relevant to Moral Epistemology?

The Use of CSM to Undermine Moral Judgments: If you believe that X is an unreliable way of forming moral judgments, and CSM reveals to you that some of your moral judgments were arrived at via X, this will lead you to doubt those judgments.

A possible example: If you believe that disgust is an unreliable guide to immorality and the CSM (in the form of Haidt's social intuitionism) convinces you that you think sibling incest is immoral just because you find it disgusting, this will lead you to doubt your prior judgment that sibling incest is (invariably) immoral.

Notice that some non-inferential sources of belief do seem to be reliable. Think, for example, of sense perception. If I say "The sky is blue," you don't undermine my belief by showing me that I "only" think this because the sky looks blue to me. In a similar vein, one might think that some non-inferential sources of moral judgment are not "debunking" in the way that disgust arguably is. (Even the realization that a moral judgment has its source in disgust alone is not debunking for those religious conservatives who think that disgust is a reliable guide to the immorality of an action. An explanation of the origins of a person's belief only debunk that belief is she *also* believes that the mechanism or process cited by the explanation is an unreliable one.

An Illustration: Suppose I believe that it's wrong to use people as slaves or servants or tools without consideration of their interests, goals and needs. And suppose I cannot defend this belief

with an argument. If the cognitive science of morality reveals that I "only" have this intuition because I empathize with people in general or have compassion for people who are being used to serve the interests of others, it is not clear that this will undermine my belief. If I believe that empathy and compassion can be components of a reliable source of moral reasoning, the CSM explanation of my belief in the immorality of using people as mere means won't undermine that belief.

Question: Does this blunt the skeptical feel of Haidt's social intuitionism? Does it suggest that the disgust or moral condemnation he elicits is an atypical source of moral beliefs? How many of your moral beliefs are vulnerable to debunking?

Preliminary Conclusion Based on these Reflections: Insofar as the CSM sheds light on the [person-level, psychological] processes that produce our moral judgments it seems straightforwardly relevant to moral epistemology when it is conjoined with various moral/epistemic beliefs in the reliability or unreliability of various possible sources of moral judgment or intuition.

Question: But what about **neuroscience**?

"Insofar as we can infer what psychological process is being implemented by a given neural process, neuroscience can indirectly inform us about the reliability of moral judgments. But, unless we can make such inferences, how can we determine whether a neural process is likely to issue in accurate moral judgments? **We could only evaluate the reliability of a neural process by observing what moral judgments result from it, and then using armchair methods to directly evaluate the judgments themselves. This is likely to require controversial commitments to substantive normative claims.** It risks circularity, if we end up justifying a set of moral judgments by appeal to the reliability of a neural process that has been certified as reliable precisely because it has produced the relevant judgments. One could perhaps try to avoid such circularity by arguing that certain controversial moral judgments were produced by a reliable neural process because that process also reliably produces judgments that we all agree to be correct—though it is not obvious why the reliability in the latter context must carry over to the former (Kahane, 2016). But in any event, **the work neuroscience does on this approach is pretty minimal: identifying types of processes that we then try to correlate with patterns of moral judgments. What exactly is involved in these processes, at the neural level, is irrelevant.**"

3. On the Relevance of Neuroscience to Epistemic Justification

You are justified in holding some of your beliefs because you inferred what you believe in a rational, cogent or "good" way from other things you rationally or justifiably believe. But what about your belief in the premises of the relevant inference?

Reliabilism: You are prima facie justified in believing such a "basic" (or not-inferred) premise just in case your belief in it was generated by a reliable process.

Two notes: (1) Examples of reliable processes: perception, memory, counting, conceptual understanding. (2) We use "prima facie" here because the justification with which you hold the belief in question might be undermined by other factors. E.g., you might see a strange looking rabbit and be justified in believing what you see until you are approached by an otherwise reliable person who misleads you into thinking that you are hallucinating.

The Generality Problem for Reliabilism: *Which* way of specifying "the" process leading up to a judgment is the relevant one for evaluating the judgment's justification?

An example: Suppose I look across a field at dusk and come to believe on this basis that a horse is approaching. If we specify the process that generates my belief as "vision" we might conclude from reliabilism that I am justified in this belief as it was generated by vision and vision is generally reliable. If we specify the process as "identifying an object across a great distance in low light" we might conclude from reliabilism that I am not justified in this belief because identifying an object at a great distance in low light is not reliable. (Sometimes I mistake a horse for a donkey in such conditions.) Which description should we use?

D&K: No matter how we answer this question, the process must be specified in psychological terms. How a psychological process is physically implemented seems intuitively irrelevant to epistemology.

The Neuroscientist's response: But mightn't we infer the psychology from the neurology as Greene infers that deontic judgments are generated by preponent emotion-responses on the basis of some prior identification of the functions typically executed by the brain regions in question?

Problems: (a) **multiple realizability**: "many (if not all) types of psychological processes …are realized by distinct neural arrangements in different individuals, or even within the same individual over time." (b) **multiple functionality**: "any given brain area or network is likely to be involved in many distinct psychological processes in different contexts (see e.g. Pessoa, 2013)." (c) **moral relevance is not a neurological kind**: "many epistemically important distinctions that are salient when judgment-forming processes are described in psychological terms are unlikely to carve distinctions that are significant at the neural level."

4. Three Approaches to the Cognitive Science of Morality

A. Universal Moral Grammar (Mikhail): our core moral judgments reflect the working of a "moral organ": unconscious computations map input about causation, intention and action onto innately represented moral principles to produce universally shared intuitions about moral permissibility and wrongness.

B. Social Intuitionism (Haidt): See Handout on Haidt

C. The Dual Process Model of Moral Judgment (Greene): "Like Haidt, Greene holds that a great deal of moral judgment is shaped by immediate "alarm-bell"-like emotional reactions (a "system-1" type process), and that much of the justification offered in support of our moral views—including much of the theorizing of moral philosophers—is merely ex-post rationalization. But Greene also claims to find evidence for an important exception, arguing that utilitarian judgments are uniquely based in explicit reasoning (a "system-2" type process). And Greene has famously argued that this difference—which he traces to distinct neural structures—supports a normative argument favoring utilitarianism."

5. On the Domain Specificity of Moral Cognition and Its Relevance to Moral Epistemology

Scientific Question: When we form moral intuitions and judgments, are we utilizing a capacity that is specific to morality—or even a "moral module"—or are we merely drawing on general psychological capacities? Relatedly, are there specific brain areas dedicated to moral cognition?... The evidence so far has not been very kind to the idea of a dedicated moral module, at least not

one with unique neural correlates.

Our Best Current View of the Neural Correlate of Moral Cogntion: moral cognition employs many neural networks involving areas distributed around the brain, including the ventromedial and dorsomedial prefrontal cortex (vmPFC and dmPFC), the temporoparietal junction (TPJ), the precuneus, the posterior cingulate cortex (PCC), the amygdala, and the temporal pole. Moreover, these brain areas are not used exclusively for moral cognition.

A Recurring Question: **Doesn't this depend on how we individuate or define moral cognition?** (See Handout on Stich.) Are debates about how to best define "moral cognition" are conceptually connected to questions about which kinds of actions, people and institutions are immoral and which are not? If so, how can we limit our normative assumptions for the sake of the kind of neutrality that many of us think of as a hallmark of science?

Returning to our Epistemological Question: Are the processes that generate our moral beliefs reliable? D&K: If moral cognition is domain general, it's harder to give a negative answer to this question. To argue for the unreliability of moral judgment one would either have to embrace wholesale skepticism about the justification of belief (which would undermine our confidence in the science too) or one would have to argue that generally reliable ways of arriving at beliefs are unreliable when used to render judgment on moral matters. One way of pursuing the latter strategy would be to argue against moral ratonalists (and Parfit) that emotions play a central role in moral judgment. But this assumes that we have independent evidence that emotions detract from the reliability of judgments rendered in other (non-moral) domains. And D&K rightfully cast doubt on the view that emotions are typically distorting factors. (Doesn't empathy often make our moral judgments **more** reliable?)

A Problem with the Attempt to Undercut the Reliability of Moral Judgments on the Basis of their (Arguably) Essentially Emotion-Laden Character: "In the mid-20th century, it was thought that emotional and cognitive processes were supported by distinct, dedicated brain regions (the "limbic system" and the neocortex, respectively). However, there is growing consensus that there is no such sharp anatomical divide in the highly interconnected human brain (e.g. Barrett and Satpute, 2013; LeDoux, 2012; Lindquist and Barrett, 2012; Okon-Singer et al., 2015; Pessoa, 2008). Brain areas involved in emotion play a crucial role in cognitive functions such as learning, attention, and decision-making, and emotions depend on brain regions involved in various cognitive operations. Consequently, cognitive scientists increasingly reject the characterization of psychological processes involved in moral judgment as either "emotional" or "cognitive" (e.g. Cushman, 2013; Huebner, 2015; Moll, De Oliveira-Souza and Zahn, 2008). If emotions are not easily separable from the "cognitive" means by which we represent, process, and evaluate the world, then this opens up the possibility that emotional processes—not just "cold" reasoning—can make rational, evidence-sensitive contributions to moral judgment.

This suggestion is supported by recent research on the network of brain areas that Greene associated with "emotional" moral judgments – in particular, the right temporoparietal junction (rTPJ), the amygdala, and the ventromedial prefrontal cortex (vmPFC)…. it seems as if "emotional" processes, mediated by areas such as the rTPJ, amygdala, and vmPFC, allow us to process morally relevant information and thus can contribute to good moral reasoning. This is supported by studies of clinical populations with neural abnormalities that disrupt these circuits, such as psychopaths, patients with vmPFC damage, and patients with behavioral-variant fronto-temporal dementia (FTD). Whilst much of their ability to reason remains intact, these clinical populations suffer from emotional deficits, including diminished empathy, that restrict their "cognitive" abilities to correctly perceive, attend to, and take into account morally-relevant

properties. For example, patients with vmPFC lesions and FTD patients have an impaired ability to infer which emotional states others are experiencing (Shamay-Tsoory and Aharon-Peretz, 2007); vmPFC patients struggle with decision-making because they lack appropriate emotional reactions (Damasio, 1994) and they fail to respond to harmful intentions (e.g. judging that attempted murder is permissible; Young, Bechara et al., 2010); and FTD patients display sociopathic tendencies. Psychopaths display abnormal functioning in the amygdala, vmPFC, and TPJ when presented with moral transgressions, and consequently fail to attend to and correctly process morally salient properties such as harm and mental states (Decety et al., 2015; Harenski, Harenski et al., 2010; Hoppenbrouwers, Bulten and Brazil, 2016)…. **These clinical populations also give abnormally high rates of so-called "utilitarian" judgments that it is morally permissible to sacrifice one person as a means to saving others. Rather than being the product of good moral reasoning unfettered by irrational emotion, it seems these judgments are associated with these patients' failure to register complex facets of moral value** (Gleichgerrcht et al., 2011)…. **Indeed, in non-clinical populations, so-called "utilitarian" judgments that Greene associated with the dorsolateral prefrontal cortex (dlPFC) are not associated with genuinely utilitarian, impartial concern for others, but rather with rational egoism, endorsement of clear ethical transgressions, and lower levels of altruism and identification with humanity** (Kahane et al., 2015). Furthermore, **in a study by Feldman Hall and colleagues (2012), where participants decided whether to give up money to stop someone receiving painful electric shocks, activity in the dlPFC was associated with self-interested decisions and decreased empathic concern, whilst the vmPFC was associated with prosocial decisions**. Moreover, the "emotional" circuits seem to facilitate truly impartial, altruistic behaviour. For example, Marsh and colleagues (2014) found that extraordinary altruists have relatively enlarged right amygdala that are more active in response to other people's emotions—which they are *better* at identifying."

6. Moral Nativism and Moral Learning

Moral Nativism: a full account of human moral psychology will need to make significant reference to features of our minds that are organized in advance of experience.

According to non-nativists, by contrast, moral judgments are best explained as the product of learning processes interacting with the environment, and little reference to specialized innate structure is needed.

Much work in the CSM has been dominated by strong nativist assumptions. UMG theorists claim that moral judgment is produced by an innate moral module while both the Social Intuitionist and Dual Process Models explain patterns of moral intuitions by reference to emotional responses selected by evolution. However, evidence for these claims is fairly limited.

The Relevance of Innateness to Moral Epistemology: Debates about the evolutionary (or other) sources of moral judgment are obviously of great interest, but their epistemic significance isn't straightforward. UMG theorists occasionally write as if the aim of moral philosophy is to uncover the innate moral code posited by UMG theory but this is an odd idea. If this innate moral code is the product of natural selection, why should we let it guide our actions? After all, evolution "aims" at reproductive fitness, not at moral truth. Why think that dispositions that were reproductively advantageous to our ancestors in the savannah track any kind of moral truth? These kinds of considerations lead Greene (2008) to a contrary conclusion: if certain moral judgments have their source in our evolutionary history then they should be treated with suspicion. Instead, we should use our general capacity for reason to arrive at independent, consequentialist conclusions.

Question: Does your belief in the innateness of a moral principle or the innateness of those faculties you employed when coming to accept that principle **undercut** or **further entrench** your confidence in that principle? Remember how Darwin answered this question?