MORAL PERCEPTION AND THE RELIABILITY CHALLENGE

David Faraci

Moral perception has received a good deal of recent attention. There are a number of motivations for this, for instance certain phenomenological similarities between moral and perceptual experiences. But one distinctively *epistemological* motivation stands out. Given a traditional intuitionist moral epistemology, it is notoriously difficult for realists to account for even the *possibility* of moral knowledge. This has led some realists to go looking for an alternative to intuitionism. Perception is an obvious contender.

If this is one's motivation, however, it is not enough to defend the existence of moral perception. One needs to show that perceptualism places one in a better epistemological position than intuitionism. One cause for worry here is that on many accounts moral perception is dependent on *a priori* background moral knowledge. While such accounts might enjoy some advantages over intuitionism, they are clearly no better off where certain fundamental epistemological challenges are concerned. In particular, such forms of perceptualism enjoy no advantage when it comes to arguably the *most* fundamental epistemological challenge to moral realism: the challenge to *explain the reliability* of our moral beliefs ('the reliability challenge').[1]
I argued elsewhere that *all* perceptualists have this problem, that moral perception is necessarily dependent on non-perceptual background knowledge of moral bridge principles (Faraci 2015).[2] In a recent response, Preston Werner (forthcoming) contends that my argument rests on an over-intellectualized picture of perception. In this paper, I argue that though Werner may well be correct, my arguments, properly

---

[1] Or at least show that their commitments do not rule out the existence of such an explanation. This is sometimes put instead in terms of explaining our *access to* the moral truth, or the *correlation between* our moral beliefs and the moral truth (Enoch 2011, chap. 7). Like Enoch and many others, I also take popular genealogical challenges (e.g., Street's (2006) evolutionary debunking argument) to be a close relatives of the reliability challenge.
To be clear, this is a challenge *for epistemologists*, a challenge for accounts of moral knowledge *per se,* not a challenge to individuals' moral knowledge. Nearly everyone agrees that we can know things without being able to explain our reliability (e.g., ancient Greeks knew things by perception). Some take there to be a derivative challenge for individuals: if we discover that there is no explanation for our supposed reliability in a domain, that serves as a defeater for our beliefs in that domain (e.g., Field 1989). This is orthogonal to the discussion on hand.
[2] In that paper, I did not explicitly frame things in terms of the reliability challenge, but recent work in this area, including Werner's, has helped me to recognize this as my central concern (and I think I can charitably be read as implicitly focusing on this—see, especially, fn. 13 in that paper).

extended, still suggest that perceptualism leaves realists in no better position than intuitionism when it comes to the reliability challenge.[3]

1. My Previous Argument

The following is a reconstruction of my (2015) argument (with some minor changes to avoid snags irrelevant to the current discussion[4]):

> **P1** Moral experiences are *inferred* (in the broadest sense, which includes sub-personal processing[5]) from non-moral experiences.
>
> **P2** If A's moral experience of M produces knowledge, where M is inferred from her non-moral experience of N, A must know that N implies M.
>
> **P3** Knowledge of *moral bridge principles* such as *N implies M* is dependent on non-perceptual moral knowledge.
>
> **C** Therefore, perceptual moral knowledge is dependent on non-perceptual moral knowledge.

In defense of P1: Adapting the famous example from Harman (1977), when we see wrongness in the burning of a cat, that experience is inferred (again, in the broadest sense) from our experience of the descriptive features of cat-burning (a description I'll spare you). Few doubt this (Werner does not), but for those who do, I argued via a counterfactual test: the 'Convincing Fake Test'. Suppose we were to construct a convincing but fake cat-burning, one that apparently has the same descriptive features as an actual cat-burning (animatronic cat, holographic fire, etc.), but which lacks any wrong-making features. Insofar as someone tends to perceive wrongness in cat-burnings, they would surely tend to experience such convincing fakes as wrong. The best explanation for this tendency is that the original experience of wrongness was dependent on the experience of the relevant descriptive features.

---

[3] For arguments related to both my earlier arguments and those raised herein, especially regarding the parity between perceptualist and intuitionist explanations for the reliability of our moral judgements, see Crow (2016) (though note that Crow limits his target to non-naturalism).

[4] Thanks to thank Preston Werner for alerting me to some of those snags, both in print and in discussion.

[5] Some may object to the use of 'inference' in referring to sub-personal processing. In this context, at least, nothing hangs on the choice of terminology. Those who prefer other usage can mentally substitute their preferred term. All that matters here is that there is a 'transition' from non-moral to moral experiences (or, perhaps better, a 'dependence' of one on the other, as the 'transition' may be instantaneous). As acknowledged in my earlier work, this requires the assumption that non-moral and moral experiences have different contents. This is largely irrelevant here, though I address it briefly in §3.

In defense of P2: Suppose Sam experiences a cat-burning. And suppose he infers that what he experiences is wrong. It is hard to see how Sam could know that what he experiences is wrong if he *doesn't* know the inference-grounding implication from cat-burning to wrongness. More generally, it seems plausible that inferences only produce knowledge if the inference-grounding implication is known.[6]

In defense of P3: Suppose Sam knows that cat-burning implies wrongness *by perception.* Surely, perception of a principle depends on perception of the relevant *relata*; principles are not the sort of thing we can directly perceive. Consider, for instance, knowledge that when the air smells a certain way, it will (likely) rain soon. We experience the air's smelling a certain way. And we experience rain. And we notice a correlation between the two. Arguably, we can thereby come to know, by perception, that a certain air quality implies rain.

Suppose this is how Sam came to know that cat-burning implies wrongness. He had numerous experiences of cat-burning (yikes!) and of wrongness, and noticed a correlation between the two, which is why he believes that the former implies the latter. If we accept P1 and P2, it follows that Sam's experiences of wrongness are dependent on background moral knowledge, so there must be some *further* principle Sam is relying on, such as that *causing suffering* implies wrongness. How does Sam know that causing suffering implies wrongness? Well, if it is by perception, then there will have to be some *further* principle that grounds his experiences of wrongness. And so on. The regress must stop at some point, and it will stop with a non-perceptual ground for Sam's belief.

In brief, then, my argument was that we infer moral experiences from non-moral experiences (P1); such inferences can only produce knowledge if we know the relevant implication (P2); and knowledge of such implications can't be perceptual, on

---

[6] This is not wholly uncontroversial, though I find supposed counterexamples (see, e.g., Fitelson 2010; Warfield 2005) unconvincing—something I cannot fully address here. What matters for our purposes is that even if the general principle is false, purported counterexamples all have certain features that Sam's and other relevant cases lack. Note also that to the extent the principle *is* correct, it seems clear that it really is *knowledge* that is required; even justified true belief won't do the trick. Suppose Sam justifiably, but falsely, believes that the amount of disutility created by burning a cat always outweighs the amount of utility sadists get from it. Sam is also justified in believing, falsely, that utilitarianism is true. Sam therefore believes that cat-burning is wrong, and infers from this that the particular cat-burning in front of him is wrong. And he's correct: cat-burning *is* wrong, just not for the reasons he thinks, so Sam doesn't know that cat-burning is wrong; he's been Gettiered. Surely, he therefore also doesn't know that *this particular* cat-burning is wrong, given that he inferred the latter from the former.

pain of regress (P3). It follows that if there is perceptual moral knowledge, it is dependent on non-perceptual knowledge of moral bridge principles (C).[7]

2. Werner's Response

Werner objects to P2. He more or less accepts that things are as I suggested when it comes to explicit, person-level inferences. But he argues that things are different in the case of *sub-personal* inferences. For there are plenty of cases, Werner maintains, where one experience is sub-personally inferred from another, and this produces knowledge, yet we have no *beliefs* about the relevant implication, and therefore no knowledge. For example:

> I can, typically, effortlessly distinguish the sound of a piano from that of an acoustic guitar. But I couldn't even begin to explain this difference or point to the low level qualities of tone and timbre that ground their differences. I have no beliefs, much less knowledge, of how I go from low level auditory information to the auditory experience of a piano. (Werner forthcoming)

Werner therefore proposes that we replace P2 with the weaker:

> **P2**[*]   If A's moral experience of M produces knowledge, where M is inferred from her non-moral experience of N, then either (a) A knows that N implies M or (b) A's "perceptual system contains subdoxastic information states which ground reliable transitions from perceptual information about [N] to perceptual information as of [M]" (Werner forthcoming)

---

[7] There is an obvious worry here that this argument will overgeneralize, suggesting that *no* (or very little) knowledge is independently perceptual. For instance, bus-perception is clearly dependent on perception of certain shapes, colors, etc. Does this mean that we need to know that *shapes and colors like that* imply *bus?* There are really two issues here. The first anticipates Werner's objection: I might be able to reliably infer busses from shapes and colors without *believing*—and therefore not knowing—any such implication. This will be addressed in what follows. Setting that aside, one might still worry that my argument implausibly suggests that there are no direct perceptions of busses. I partially addressed this in my (2015) by adding a caveat for experiences with the same contents; perhaps experiences of certain shapes and colors *just are* experiences of busses. There are other ways of cashing this out, and I'll remain largely noncommittal here. The core point will be that there is a distinctive problem for *moral* perception, which is that on most metaethical views, what is mysterious is our (perceptual systems') ability to learn which descriptive features imply which moral ones. By comparison, we should be able to tell a relatively simple story about how we or our perceptual systems learn to identify busses in light of their shape and color.

Consider Diane, who is just like Sam, except that instead of making an explicit inference, Diane experiences the cat-burning as wrong because she has a "subdoxastic information state" (from here I'll drop "information") that grounds reliable sub-personal inferences from experiences of cat-burnings to experiences of wrongness. It seems perfectly possible that Diane thereby knows that this particular cat-burning is wrong. Yet, as Werner points out, Diane doesn't need to *know* that cat-burning implies wrongness because Diane doesn't even need to *believe* this.

From here, I assume that Werner is correct about all this, that there may be moral perceptions that are independent of any *a priori* moral knowledge. My concern is what the move to P2* tells us about perceptualism's ability to address the reliability challenge. What it tells us, I submit, is not that perceptualism has an answer to this challenge, but only that some perceptualists face a version of the challenge that is slightly different from the one intuitionists face. If Sam explicitly infers wrongness from cat-burning because he intuits that cat-burning implies wrongness, the intuitionist needs to show that her view can explain, or at least does not render inexplicable, the reliability of Sam's moral intuition. Similarly, if Diane's perceptual system sub-personally infers wrongness from cat-burning via a subdoxastic state that grounds such inferences, the perceptualist needs to show that his view can explain, or at least does not render inexplicable, the reliability of Diane's subdoxastic state.

One might try to resist this parallel, maintaining that the standards for subdoxastic states should be different than for beliefs. But this is implausible. Suppose Sam's intuition is merely a product of his culture. (I assume merely for illustrative purposes that such an explanation has skeptical implications.) Sam therefore believes, but does not know, that the particular cat-burning before him is wrong. Now suppose that Diane's subdoxastic state has a precisely analogous cultural explanation, perhaps even the same one; it just happened to produce a subdoxastic state in her, but an intuition in Sam. Is it possible that Diane knows that the cat-burning before her is wrong? Surely not. Diane cannot possess knowledge, while Sam lacks it, simply because her inference is sub-personal, while his is explicit.

This recommends a simple expansion of the reliability challenge. Call something an *inference-grounding state* just in case it is either a belief or a subdoxastic state that grounds inferences. We should accept:

> **Reliability Challenge**   Realists must explain, or at last show
> that their view does not render inexplicable, the reliability of moral
> inference-grounding states.

I have already argued that perceptualism enjoys no advantage here when it comes to moral perceptions that are mediated by background beliefs. The question given Werner's arguments is whether perceptualism enjoys any advantage when it

comes to moral perceptions that are mediated by subdoxastic states. As I now argue, the same considerations that led me to deny perceptualism any advantage in the former case likewise support denying it any in the latter.

3. Moral Perception and the Reliability Challenge

As seen in §1, I previously offered a regress argument against the possibility of independent perceptual moral knowledge. Framed in terms of the Reliability Challenge, the upshot of my argument was meant to be that accounting for the reliability of perceptual moral beliefs will always require accounting for the reliability of non-perceptual moral beliefs.

We might simply generalize my point in light of Werner's arguments; the claim would be that accounting for the reliability of perceptual moral beliefs requires accounting for the reliability of non-perceptual subdoxastic states. The problem with this is that we don't have a good account of what makes a subdoxastic state perceptual or non-perceptual. By definition, if an inference is dependent on a non-perceptual belief, the inferred belief is not independently perceptual. But the same is not true of sub-personal inferences: even if a subdoxastic state is not formed by perception, the inferences it grounds might still count as independently perceptual. Suppose, for instance, that we are evolutionarily hard-wired to experience certain facial expressions as expressing certain emotions—e.g., we possess a subdoxastic state that grounds sub-personal inferences from (say) experiences of scowls to experiences of anger. Arguably, my knowledge that my interlocutor is angry might still count as independently perceptual.

This suggests that I was mistaken to frame things in terms of whether moral epistemology is "purely" perceptual (Faraci 2015, 2059). But it does not suggest that I am wrong to think that perceptualists enjoy no advantages in explaining our reliability. For our guiding question now is not whether the relevant subdoxastic states are in some sense perceptual, but whether available accounts of their reliability, or evidence that such an account is compatible with realism, *rely* on their perceptual nature. If not—if the accounts on offer are the same or sufficiently similar to the accounts an intuitionist might give—then perceptualism isn't any better off than intuitionism where the Reliability Challenge is concerned.

The problem for perceptualism here is simple: perception alone *never* provides an answer to the Reliability Challenge. We always need to account for the reliability of perceptual experience. The reason perception can *seem* promising is that the reliability of a great deal of perceptual experience seems fairly easy to account for. Consider, for example, Werner's discussion of chair-perception:

> We presumably don't have an innate representation of chairs. Nor do
> we learn what chairs are by *a priori* reflection. Rather, we have an

ability to gain the ability to perceive chairs via a complicated feedback loop of reinforced and undermined perceptions of chair-like objects. (Werner forthcoming)

This is an attractively simple story, one that seems well-suited to account for our reliability. First, there is a tight connection (perhaps identity) between the bits of the world that explain our chair-experiences and *being a chair*. Second, in part because we gain competence with the concept CHAIR through the relevant feedback process, there is a similarly tight connection between those bits of the world and our conceptual model of a chair. Putting these together, it might be that reliably identifying chairs requires nothing more than consistently applying our CHAIR concept.

Contrast this with perception of H_2O. As with chairs, there is a tight connection (perhaps identity) between the bits of the world that explain my H_2O-experiences (water) and *being H_2O*. But unlike with chairs, we do not perceive something as H_2O merely by applying our concepts of HYDROGEN, TWO and OXYGEN. To explain our reliability, therefore, we need to tell a further story about how we learned to identify the relevant bits of the world as H_2O.

Which of these stories the moral realist can echo depends on her other metaethical commitments. To echo the chair-perception story, the realist must hold that Diane can reliably identify wrongness merely through consistent application of her concept WRONG. Arguably, such an account is available only to analytic naturalists.[8]

Other realists can tell a story echoing the one about H_2O-perception, on the assumption that the supervenience relation between (e.g.) wrongness and cat-burning can play the same role here as the identity relation between water and H_2O. Beliefs that cat-burnings are wrong, inferred from experiences of cat-burning, reliably track wrongness because wrongness supervenes on cat-burning (we may suppose). But as with H_2O-perception, these perceptualists need an account of how we come to reliably identify cat-burnings *as wrong.*

It is here that my regress rears its head. Diane's perceptual system must *learn* to identify cat-burnings as wrong. As I pointed out, if her system learned this through experience, it seems this could only have been by correlating independent experiences of wrongness and experiences of (features of) cat-burnings. Given the

---

[8] I previously acknowledged that of all realists, analytic naturalists are most likely to be able to sidestep my arguments (Faraci 2015, §2.3). Such theorists might hold, I suggested, that moral and non-moral experiences have the same experiential content, and thus that no background knowledge is needed to 'transition' between non-moral and moral experiences. If seeing something *as cat-burning* is experientially the same as seeing it *as wrong,* then arguably no background knowledge is needed to facilitate an inference from one to the other.

inferential nature of moral experience (as per P1 of the argument in §1), those independent experiences of wrongness would have to be dependent on her perceptual system's having learned *further* such correlations, which requires further moral experiences. And so on.[9] The relevant lesson is that the explanation for Diane's ability to reliably identify cat-burnings as wrong will have to appeal to something antecedent to the workings of her perceptual system.

Of course, that the solution to the Reliability Challenge must be antecedent to the workings of our perceptual systems doesn't suggest that perceptualists can't meet that challenge. But it *does* suggest that it will not be the perceptual elements of the view that provide the solution to the challenge, and thus that the solution, whatever it is, will not be *distinctively perceptualist-friendly.*

Of course, we can't be sure that this is the case until we know how perceptualists will actually meet the Reliability Challenge. But looking at Werner's gestures in this direction help bolster the point. Start with the analogy with chair-perception. Suppose someone held, contra Werner, that our concept CHAIR is innate, and thus that chair-perception is facilitated by *a priori* knowledge. Is this intuitionist view any worse off where the Reliability Challenge is concerned? It is hard to see how. Surely it is the fact that being reliable requires only conceptual competence, not the fact that such competence is gained perceptually, that accounts for this view's ability to meet the Reliability Challenge. Likewise, it would seem to be the *analytic* not the *perceptualism* in 'analytic perceptualism' that does the heavy lifting where the Reliability Challenge is concerned. If any intuitionists can answer that challenge, it is analytic naturalists.

Next consider two suggestions Werner gives concerning the potential workings of our moral perceptual systems. He says that perhaps we can capitalize on our affective system's "ability to represent objects, people, and events as aversive or attractive. Or perhaps the perceptual system alone has evolved to represent things in this way, in order to facilitate quick action" (Werner forthcoming).

True, our affective systems can represent things as aversive or attractive. But this feature of our affective systems has often been used to *challenge* our reliability, not vindicate it. It is sometimes argued that certain moral reactions' affective provenance should lead us to shun them in favor of ones with more 'rational' bases.[10]

---

[9] By contrast, the story about H_2O-perception can stop after a finite number of iterations. Someone might correlate independent experiences of water with hydrogen and oxygen molecules viewed through an electron microscope. Then we may have to consider how he learned to identify what he sees in the microscope *as* hyrdrogen and oxygen. Unlike in the moral case, there is no principled reason to think this regress will continue indefinitely.

[10] See, e.g., Greene (2008) and Singer (2005). To be clear, I am not endorsing such arguments; I agree more with opponents such as Kahane (2014).

The implicit assumption here is that the former are less likely to be reliable, that it is our reason, rather than our affective responses, that reflects the moral truth. We can see why replying that such affective responses are perceptual is attractive; calling something perceptual implies that it *does* reflect the truth. But without an account of how, or at least evidence that this is the case, the Reliability Challenge goes untouched, and the appeal to perception is merely a distraction.

Similarly, it is hardly seen as an epistemic boon for our moral beliefs that we have evolved to have them. This, too, stands at the center of a popular epistemological objection: since there is no apparent reason to believe that evolution tracks the moral truth, the evolutionary origins of our moral intuitions should make us doubt their reliability.[11] It is hard to see why the lesson changes if we shift from an explanation of moral beliefs to an explanation of moral experiences. An evolutionary explanation seems no better or worse suited to explain the reliability of moral perception than it does moral intuition.

4. Conclusion

Perceptual epistemology is attractive because for much of our perceptual knowledge, our reliability does not seem particularly mysterious. Perhaps I know that's a chair simply because the physical object impedes upon my experience and fits my concept CHAIR. Analytic naturalists may be able to tell an equally simple and compelling story about moral knowledge. But it is their analytic naturalism, not their perceptualism, that gives them this ability. What remains for everyone else is a puzzle about how our perceptual systems learn that certain non-moral experiences are indicative of certain moral features. That puzzle is only marginally different from, and arguably no less difficult than, the analogous puzzle regarding moral intuitions. There may well be many other reasons to pursue moral perception—even, as Werner has shown, epistemologies that are in an important sense purely perceptual. But perceptualism ultimately does nothing to address arguably the deepest worry about the possibility of moral knowledge.

Acknowledgements

References

Crow, D. 2016. "The Mystery of Moral Perception." *Journal of Moral Philosophy* 13 (2): 187–210.

---

[11] The most popular incarnation at present is Street (2006).

Enoch, D. 2011. *Taking Morality Seriously: A Defense of Robust Realism.* Oxford
        University Press.
Faraci, D. 2015. "A Hard Look at Moral Perception." *Philosophical Studies*, no. 172:
        2055–2072.
Field, H. 1989. *Realism, Mathematics, and Modality.* Oxford, UK ; New York, NY,
        USA: Blackwell.
Fitelson, B. 2010. "Strengthening the Case for Knowledge from Falsehood." *Analysis*
        70 (4): 666–69.
Greene, J. 2008. "The Secret Joke of Kant's Soul." In *Moral Psychology: The
        Neuroscience of Morality*, edited by W. Sinnott-Armstrong, 35–79. MIT Press.
Harman, G. 1977. *The Nature of Morality: An Introduction to Ethics.* Oxford
        University Press, USA.
Kahane, G. 2014. "Intuitive and Counterintuitive Morality." In *Moral Psychology &
        Moral Agency: Philosophical Essays on the Science of Ethics*, edited by J.
        D'Arms and D. Jacobson, 9–39. Oxford University Press.
Singer, P. 2005. "Ethics and Intuitions." *Journal of Ethics* 9: 331–52.
Street, S. 2006. "A Darwinian Dilemma for Realist Theories of Value." *Philosophical
        Studies* 127 (1): 109–166.
Warfield, T. 2005. "Knowledge from Falsehood." *Philosophical Perspectives* 19: 405–
        16.
Werner, P. J. forthcoming. "Moral Perception Without (Prior) Moral Knowledge."
        Journal of Moral Philosophy. https://doi.org/10.1163/17455243-46810075.