

This essay is split into two parts. In part 1 I argue that a moral intuition is an attitude, which is differentiated from other attitudes phenomenologically by 'shouldness', I also explain how, contra Blackburn, moral intuitions are not to be identified functionally. In part 2 I explain how trusting an intuition, or attitude, works, and which attitudes survive filtering. Part 1 assumes sentimentalism, part 2 relies on anti-realism, though I will not argue for anti-realism as opposed to realism. Part 2 also assumes expressivism in Blackburn's quasi-realist form, that is, the view that moral judgements express our attitudes rather than report facts, and that truth talk about them is earned from the attitudes themselves rather than read off the world. In light of this, one might accept part 1, but disagree with part 2, and so disagree with my conclusion. One might ask whether "should we trust our intuitions?" is itself a moral question, in which case, on my own account, my answer would have no truth value. I understand the should as that of warranted reliance, as in "should I trust my memory", not the should of "should I lie".

I claim that a moral intuition is a first-order affective attitude ¹ and will justify this with the following arguments:

First is the observation that our moral intuitions can be recalcitrant. A moral seeming can persist against one's contrary beliefs. Huckleberry Finn believes he ought to turn Jim in, yet it does not seem right to him, and he acts on the seeming². One might object that beliefs too can persist against evidence. But believing that P while believing that not-P is straightforwardly irrational, whilst feeling that P while believing that not-P is not irrational.

Second, if we introspect at a moment we have a live moral intuition, the phenomenological character we experience is that of an emotion, not a neutral feeling of holding a belief. One might argue that beliefs do the moral work and our attitudes are by-products, but the feeling itself presents acts as moral or immoral, that is, the affect is the mode in which morality appears, and so the content cannot be separated from the emotion. For example, indignation at a queue-jumper is not a neutral pang accompanied by a separate judgement that the act is unfair, the feeling itself presents the act as unfair.

Hence moral intuition behaves like an attitude and not a belief. Since a seeming can sit alongside the contrary belief without contradiction, its felt character is that of an emotion, and emotions are attitudes. Since moral intuitions are attitudes, I will speak of attitudes, or of first-order moral attitudes, from here on.

I claim moral attitudes are marked off from mere preferences and other attitudes phenomenologically, by a felt shouldness, which is a positive, neutral or negative feeling about whether something ought to happen. Shouldness is a primitive quality, like the quale of seeing red, or feeling pain, we identify shouldness by feeling it, not by analysis or pointing to something

¹ A. Kauppinen, 'A Humean Theory of Moral Intuition', *Canadian Journal of Philosophy* 43:3 (2013), pp. 360–381. I take the psychology from Kauppinen but not his metaethics, since he pairs it with response-dependent moral properties which intuitions can veridically present, and my anti-realism does not allow those.

² The case originates in J. Bennett, 'The Conscience of Huckleberry Finn', *Philosophy* 49 (1974), pp. 123–134.

beyond it. Hence a demand to define the feel of shouldness is as legitimate as a demand to explain what it is like to see red³.

Content does not differentiate moral attitudes from others. Consider an advanced civilisation which can control the weather. A minister decides it will rain for a month, your house floods, and you have a negative attitude towards the rain. That attitude may be a simple negative feeling, or it may carry a should not feeling. An attitude can be directed at the same content and be moral or not. One might also argue that the moral-ness arises when we consider the agent's involvement, however we frequently have attitudes towards the actions of agents, which are not moral attitudes. Hence the direction of an attitude at an agent is not sufficient for that attitude to be moral. It can be said, that agency is a necessary background to shouldness, we will not feel a shouldness if there is no agent⁴.

One prominent proposal distinguishes moral attitudes by their function, for example, they govern choice, and we demand them of others⁵. I accept Blackburn's expressivism but reject this demarcation, and the two come apart since his account of how attitudes work together does not depend on his account of which of them are moral. Each marker is routinely absent. We knowingly act against our own moral attitudes, during which the attitude stays moral through its failure to govern choice. We do not universally demand our attitudes of others, for example I might even endorse my friend doing an immoral thing because he is my friend and I favour him. It may be replied that these are dispositions, and dispositions are defeasible by nature. But consider the cold enforcer, a person who demands a norm of everyone, holds it authoritative, and lets it govern his every choice, while feeling nothing at all. He has the whole function and none of the sentiment, and his attitude is not moral. What stays fixed across these cases, present in the akratic and the partial and absent in the enforcer, is the felt shouldness. Hence, moral-ness lies in the attitude's character, not its role.

Trust understood epistemically, for example to trust intuitions to track some truth, is unavailable under anti-realism, as there are no such truths. Nor is this peculiar to intuitions since moral theory and testimony are in the same position. So, if trust is to mean something it must be in a practical sense. It may initially be unclear how one might 'trust' a moral attitude, since we tend to view trust as a rational activity. But trust is never a purely rational verdict. We may trust on evidence, but we also trust where there is no reason to, on nothing more than a feeling that something is trustworthy. We may continue to trust even when reason points the other way, consider Mitchell's parable of the partisan⁶. Hence trust is our own endorsement of our reliance

³ F. Jackson, 'Epiphenomenal Qualia', *Philosophical Quarterly* 32 (1982), pp. 127–136. Mary knows every physical fact about colour vision but learns something new on first seeing red. The parallel claim here is not that shouldness is non-physical, only that it is not conveyed by description.

⁴ Not every attitude expressible with 'should' is moral, the prudential 'you should leave now' carries a shouldness of its own. I claim only that the moral variety is distinct and that we can feel the difference.

⁵ The sort of role Blackburn identifies in *Ruling Passions* (Oxford, 1998).

⁶ B. Mitchell, 'Theology and Falsification', in A. Flew, R. M. Hare & B. Mitchell, *New Essays in Philosophical Theology* (SCM, 1955). A partisan in occupied territory meets a Stranger who tells him he is on the resistance's side and asks to be trusted. Thereafter the partisan sometimes sees him aiding the resistance and sometimes sees him in police uniform handing patriots over, and through it all he goes on

on something. Specifically in the moral case trust is a higher-order attitude that endorses reliance on a first-order attitude in moral deliberation about what to do. In this way trusting an attitude or emotion is not a category error, since trust was always an attitude of reliance rather than a proof. Higher-order endorsement is not automatically warranted, the fanatic endorses his attitudes wholeheartedly, without ever examining them. Trust is the endorsement and warranted trust is the endorsement that survives the filter.

The trap is that 'warranted' looks like a realist property with no facts to be a property of. If warrant means answerable to something outside of attitudes, then I must concede that there cannot be warranted trust for a moral attitude. There also is no escape up the order of attitudes because if trust is warranted only when a further attitude endorses it, the same question arises about that attitude.

On the quasi-realism about morality I have discussed here, truth-talk is earned from disciplined attitudes, one can say 'kicking dogs is wrong is true', not because it matches a fact, but because saying 'kicking dogs is wrong' is the expression of an attitude, and 'is true' adds nothing beyond re-asserting it. To say trust in an attitude is warranted functions similarly. To call a trust unwarranted is not to report a property it lacks but to express an attitude towards it. The 'unwarranted' attitude is disapproval of an endorsement that could not survive being examined honestly. I claim that examined honestly requires a filter with three conditions, full non-moral information, exposed framings and an honest view of where the attitude came from, that is its genealogy. There is no regress here, the attitude of warranting an endorsement is not conferred from a higher attitude since it occurs when conditions internal to endorsing are met. There is no separate act of endorsing the filter, if endorsement survives full transparency then our attitude that it is warranted follows.

The first condition of the filter is about our non-moral information. Our attitudes are directed at acts as they are construed to us, construal is fixed by our non-moral beliefs which are truth-apt. False construal can cause us to have a moral attitude aimed at something that misrepresents reality or something that does not exist. Our moral attitude must survive full non-moral information for higher-order endorsement of it to continue.

The second aspect of the filter is framing-sensitivity. Suppose I approve of a policy described as saving two hundred of six hundred, and disapprove of the same policy described as letting four hundred die⁷, I would have positive and negative attitudes to the same policy. Neither attitude is false, the clash is between attitudes, not between an attitude and the world. But recall what the attitude is for. Moral attitudes are what I rely on in deliberating about what to do, and a pair that is at once for and against the same act returns no answer. Once the effect is exposed, to go on relying on it is to let the wording decide.

trusting. Mitchell's purpose is a question about falsification in religious language, but the parable also shows trust outrunning the evidence without becoming irrational.

⁷ The example is Tversky and Kahneman's Asian disease problem, 'The Framing of Decisions and the Psychology of Choice', *Science* 211 (1981), pp. 453–458. On framing effects in moral intuition specifically, see W. Sinnott-Armstrong, 'Framing Moral Intuitions', in *Moral Psychology*, vol. 2 (MIT Press, 2008).

The third condition is genealogical. A genealogy defeats when it reveals that the attitude was produced by a process I do not, on reflection, endorse as a way of coming to have moral attitudes. That I would have held the opposite attitude with equal conviction had I been born two valleys away defeats it, because I do not endorse geography as a moral authority. That an attitude is kin-selection residue defeats it, because I do not endorse gene-sharing as a ground for how others should act⁸. But that compassion was built into me by evolution defeats nothing, because I endorse compassion whatever its source. The first-order attitude is not dissolved: the partiality still pulls, as Huck's seeming still pulled against his belief. What is withdrawn is the endorsement, not the feeling.

The views and attitudes of real fanatics are held up by false empirical claims about the out-group, and framing manipulation whilst ignoring facts that contradict their worldview. As a matter of real world psychology, if we strip the real fanatic of their scaffolding, their fanatic attitudes characteristically do not survive. The ideal fanatic, a fully informed and consistent fanatic, would survive the filter, and I concede that. But warrant here is always warrant relative to an agent's own attitudes, so his trust being warranted by his lights does not make it warranted by mine, and my disapproval of him is not refuted by his consistency. Nor does he show that the filter fails to discriminate, which is what I claimed for it, since real attitudes do fail it.

Trust understood this way is perspectival, and I should say that plainly. Two agents may both hold attitudes that are fully informed, consistent, and have survived their genealogy, and those attitudes may still conflict, with no further fact to settle it. This is a result of anti-realism.

My answer rests on three claims. First, that practical trust is the only sense of trust available under antirealism. Second, that trust is itself an attitude, so trusting an attitude is not a category error. Third, that the standard discriminates, since real attitudes fail it. If those hold, then moral attitudes which survive the filter are trustworthy in the only sense the word can still have, that is, they are defeasible, correctable, and they are the only material we have. The filtering is itself steered by attitudes, including the attitude that favours examined verdicts over raw ones. The circle is real and I accept it. Foundations were only ever available to the realist, and I have given the realism up. What matters is that the circle corrects, attitudes do get defeated inside it, even though the practice answers to nothing outside itself. Neurath's image is the honest one, we rebuild the boat at sea⁹. We should stop asking our attitudes to be mirrors of a moral reality that is not there, and trust them as what they are, not mirrors, but fuel.

⁸ S. Street, 'A Darwinian Dilemma for Realist Theories of Value', *Philosophical Studies* 127 (2006), pp. 109–166. Street's argument is directed against realism. I take the attitudes as prior, which turns the debunking argument into a filter.

⁹ O. Neurath's image, popularised as the epigraph to W. V. O. Quine, *Word and Object* (MIT Press, 1960).

