Who's Afraid of Trolleys?

Antti Kauppinen

To appear in *Methodology and Moral Philosophy*, ed. Jussi Suikkanen and Antti Kauppinen.

Routledge.

Abstract

Recent empirical studies of philosophers by Eric Schwitzgebel and others have seriously called into question whether professional ethicists have any useful expertise with thought experiments, given that their intuitions appear to be no more reliable than those of lay subjects. Drawing on such results, sceptics like Edouard Machery argue that normative ethics as it is currently practiced is deeply problematic. In this paper, I present two main arguments in defense of the standard methodology of normative ethics. First, there is strong reason to believe that expertise with thought experiments requires considering scenarios in their proper theoretical context and in parallel with other pertinent situations, so that we should not expect philosophers to be better than lay folk at responding to decontextualized cases. Second, skeptical views underestimate the epistemic benefits of the actual practices of post-processing initial verdicts both at individual and social levels. Contrary to a mythical conception of 'the method of cases', philosophers are frequently sensitive to the quality of intuitive evidence, reject and revise their verdicts on the basis of independently supported principles or interpersonal criticism, and defer to recognized specialists.

Philosophers find it hard to think about right and wrong without thinking about fanciful, even gruesome deaths – people cut up into pieces to provide organs for others, blown up to open a blocked cave, shot by a firing squad, pushed in front of a train, or ionized by death rays. (Many of these scenarios spring from the fertile, yet apparently morbid imaginations of Philippa Foot and Judith Thomson.) For some critics, the problem with this kind of approach is that it flattens the rich texture of the situations we face in real life – if we want to talk about people killed by trains, we should talk about Anna Karenina rather than obese men on footbridges. And no doubt there is much to be said for philosophical reflection on genuine ethical problems. That should be the public face of moral philosophy. Yet there is a compelling rationale for focusing on unrealistic pairs of scenarios in the specific context of

constructing theoretical accounts that aim to explain and understand the precise contours of what is morally permissible or prohibited. To isolate the features that make a moral difference, we need to examine cases that differ only along one potentially relevant dimension, but are otherwise as similar as possible.[1] And since any credible proposal for what is morally important has plausible implications for ordinary cases, we must look at extraordinary ones to find the difference.

It would evidently be pointless to consider what a principle entails about the moral status of a possible action if we didn't have some independent grasp of whether the act is right or wrong. But *can* we really tell what is right or wrong in the kind of recherché scenarios that are the lifeblood of normative ethics? There is an abundance of empirical evidence that people in general are bad at judging unfamiliar situations, and can easily be swayed by irrelevant contextual or presentational factors. It is tempting to respond, as many have done, that philosophers are special, capable of reliably zooming in on fine-grained distinctions that ordinary folk might easily miss. But this assumption, too, has recently come under attack, both in light of experiments done on philosophers themselves and in light of questions about the training that philosophers receive for the task.

My focus in this paper is on this *a posteriori* challenge to the standard methodology of normative ethics. After specifying the target a little, I'll present some of the experimental evidence and criticism based on it, drawing on the work of Eric Schwitzgebel and Edouard Machery, among others. Machery's (2017) radical claim is that given empirical data about instability and cultural relativity, we should suspend judgment about philosophical cases. The consequence is "an extensive modal ignorance that prevents us from answering

[1] Some might think that the very search for abstract, general moral principles is misguided, given the nature of the topic, or that no features are right- or wrong-making except in a particular context (Dancy 1993). I cannot address these very good worries here.

important traditional and contemporary philosophical questions" (Machery 2017, p. 185). For Machery, normative ethics as we know it is a pointless exercise.

While I agree that the empirical findings yield a prima facie reason to worry, I'm going to argue that they don't suffice to defeat the presumption that philosophers have the skills they need to make use of exotic moral scenarios. A large part of the reason why is that the experiments attempt to study an aspect of philosophical expertise *in isolation from the theoretical and practical context in which such expertise is ordinarily exercised*. There is no such thing as a 'method of cases' that could be isolated from the broader context of philosophical reflection and theorizing. But in its proper place and with appropriate (and familiar) precautions, there's good reason for normative ethicists to keep working with fanciful scenarios just as they've been doing.

1. The Setup: Why Trolleys?

Since I will be arguing that philosophical expertise with thought experiments is only exercised in its appropriate context, I'll begin by laying out some of the theoretical background of trolley cases, which is familiar to philosophers but typically ignored by psychologists doing experimental work, as Guy Kahane (2012) observes. So, consider a fundamental question in normative ethics: does it make a moral difference *how* an agent brings about an outcome? Some people, most notably act utilitarians, think it doesn't, while others think it does. How could we find out which side is right? The obvious way is to think about how the competing hypotheses fit with the things we know. This may include general truths about moral permissibility, value, rationality, and other things. But it may also include truths about particular situations, real or hypothetical (since basic moral truths presumably hold in every possible world). So part of evaluating a moral hypothesis is considering what it implies about particular cases. The act utilitarian hypothesis notoriously fares badly in this

respect, since it not only permits but requires things like killing or torturing innocents in suitable if unusual circumstances.

Recoiling from act utilitarianism, it might be tempting to think that morality prohibits doing things like killing innocent people, even if the alternative is that more people are killed. Yet this, too, appears to be false, if we consider its implications. It would mean that it is wrong to redirect a mortal threat, such as a train, toward one person, even if that meant it would crush many more people. But it is apparently not wrong to do so. At least, that is the general consensus among philosophers.[2] This suggests that doing bad things for the greater good is morally wrong only in some cases. But what is the difference between the cases in which harming some for the good of others is permissible and those in which it isn't?

It is in the context of trying to answer this question that philosophers like Philippa Foot, Judith Thomson, and Frances Kamm have introduced many variations of so-called trolley cases into the literature. Why trolleys, and sometimes trains? Presumably because the relevant moral questions are hard enough to answer even if we bracket the issues of chance and probability, and running on rails connotes a deterministic process. What is held constant in these cases – what makes something a 'trolley case' in the relevant sense – is that some greater number (usually five) of people are going to be killed by a non-intentional process (say, the movement of a runaway train, avalanche or disease), unless some agent who is not responsible for the threat intervenes in a way that somehow or another causes the death of fewer people (usually one). It is crucial that all the people involved have a stake in living on and haven't forfeited any pertinent rights, and thus have a full and equal moral status – that is, they all have an equally stringent claim on others not to harm them against their consent, regardless of whether some stand to benefit more than others or have a better character, or whatever (see e.g. Jaworska and Tannenbaum 2018). What varies between the cases is just

[2] The most notable exception being Thomson 2008.

how the death of the one person would be caused, because the point, again, is to examine which ways of bringing about the greater good are impermissible and why. The stipulation that the threat facing the people is a trolley is part of what Edouard Machery (2017, p. 13) calls the "superficial content" of the scenario, while its being a deterministic process whose direction depends on an agent is part of its "target content", the stipulations that are meant to determine which moral facts would hold in the scenario, were it real.[3]

The most famous variants are what I like to call Switch and Footbridge (Foot 1967, Thomson 1985). Switch is the case we already encountered: a trolley is heading toward five innocent people, who will be killed if a bystander does nothing. The bystander, however, can turn a switch that redirects the trolley to a side track, with the consequence that one innocent person on the side track will be killed. The bystander knows all these facts (like probability, uncertainty is bracketed here for simplicity). In Footbridge, the situation is otherwise the same, except the bystander is on a footbridge behind a large man, who would be heavy enough to stop the trolley before the five are hit, if the bystander pushed him onto its path, but this would also result in his death. (In a more carefully balanced variant, Trapdoor, the bystander could drop the large man through a trapdoor by hitting a switch some distance away.)

Most philosophers think it is morally permissible to turn the trolley in Switch, but not push the large man in Footbridge (see Bourget and Chalmers 2014). If this is right, it is evidence that some difference between Switch and Footbridge makes a difference to moral permissibility. One salient difference is that the one person in Switch is killed as a side effect of saving the five, while in Footbridge, the one person is killed as a result of being used as a means to save the five. This would fit with the Doctrine of Double Effect (DDE), the view

[3] It's worth noting that it may not be the intended target content that actually determines which facts would hold in the scenario – for example, Kamm 2015 argues that intentions are irrelevant in trolley cases, though Foot certainly thought otherwise. I'll ignore this complication in the following.

that while it is always impermissible to intend evil, an action that brings about an unintended and proportional lesser evil as a side effect of bringing about the greater good is permissible.[4] Yet this is now widely considered insufficient to explain the trolley cases. In particular, take Thomson's (1985) Loop Case: the bystander can turn the trolley away from the five onto a side track, which this time loops back towards them, and would still kill them, except that on the loop track there is a man who is sufficiently large to stop the trolley before it hits the five, though it will cost him his life. Many believe it is permissible to turn the trolley, though doing so apparently involves using the one as a means to save the five.

What exactly is the role of thinking about such scenarios in ethics? I think it is fair to say that in the standard methodology, truths about particular cases are meant to serve as an *independent check* for proposed moral principles, such as utilitarianism or DDE. If the only way to know what the right thing to do in a situation is inferring a conclusion from a principle and the non-moral features of the scenario, thinking about cases is *useless for justification*. (It might still help *understand* or illustrate the principle, and thus have a secondary epistemic role.) So for the methodology to work, there must be a way to come to know or at least form justified beliefs about right and wrong in particular cases independently of such inference.[5] The common term for non-inferential verdicts about cases is *intuition*, so we can also say that the common methodology assumes that (at least some) intuitions are trustworthy, at least to some extent.

Note that in this usage, 'intuition' neither refers to or presupposes a special faculty of intuition that allows us to grasp timeless truths. It is a much disputed metaphilosophical question just what intuitions are. Hugo Mercier and Dan Sperber offer a minimalistic metacognitive definition, observing that "When we have an intuition, we experience it as

[4] DDE was originally introduced by Aquinas to account for the permissibility of self-defense.
[5] It is, of course, also possible to form verdicts about cases on the basis of inference. The claim is just that if that's the only way, thinking about cases doesn't help in evaluating principles.

something our mind produced but without having any experience of the process of its production" (2017, p. 65). On this view, intuition contrasts with reasoning in not involving conscious inference and with perception in not being experienced as direct awareness of the world – it doesn't seem to you that you know that *p* because you see or remember it, say, but because, well, somehow you just know it (psychologists like to talk about intuition as "knowing without knowing how you know"). Along these lines, I prefer a mildly exceptionalist characterization of the relevant sense of the notion, according to which, roughly, A has an intuition that *p* if and only if A takes *p* to be the case (perhaps necessarily) if situation S obtains, without inferring this from something else, and it seems to A that the source of this verdict is some sort of insight of hers into the subject matter (in contrast to perception or memory). For my purposes here, it doesn't matter whether or not intuitions are beliefs or some other kind of verdictive state, such as a seeming that *p* (although I do happen to think I can have the intuition that *p* without believing that *p* – see Kauppinen 2013).

In contrast, some metaphilosophers claim that philosophical intuitions about cases are just ordinary judgments with no special aetiology, character, or subject matter (Williamson 2007, Machery 2017). Machery, for example, holds that "People who judge that the protagonist in a Gettier case does not know the relevant proposition deploy the capacity to recognize knowledge and distinguish it from lack of knowledge, the very capacity that allows them to judge that someone knows what she is talking about or that a karaoke singer does not know the words of a song." (2017, p. 21) And sure, people do exercise the very same competence in both cases. The difference is just that in the philosophical context, you *only* need to exercise the (in this case conceptual) competence, as long as you understand the description. To be warranted in judging whether a singer knows the words to a song, you need warrant to believe that the correct lyrics are X and warrant to believe that the singer sang something else, in addition to knowing what it is to know something. In a Gettier case,

the corresponding propositions are simply stipulated. So on the mildly exceptionalist view I hold, the key claim isn't that you need some higher sort of ability to have philosophical intuitions – it's that you need *less* than for ordinary judgment, since ordinary sources of potential error are not present, and competence and understanding suffice.[6]

Sometimes, but not always, the competence at issue is conceptual. For example, one might have the intuition that if you see a barn in a country full of indiscriminable fake barns, you don't know that there's a barn in front of you. This is something you can plausibly know simply in virtue of your competence with the concept of knowledge.[7] In the moral case, however, the Open Question Argument (Moore 1903) strongly suggests that competence with moral concepts does not suffice to tell what is right or wrong, since it seems equally competent speakers can disagree about such issues. Even if utilitarianism is false, it is not as if the utilitarian does not know what it means to say that an action is wrong. So moral competence, which includes at least a reliable disposition to discriminate between what's permissible and impermissible across a range of different cases, requires more than conceptual competence.[8]

The term 'intuition' is in some ways unfortunate, since it not only connotes a special faculty, but also suggests something like an unthinking, immediate gut reaction or hunch in everyday talk. However, as philosophers like Robert Audi (2015) emphasize, being non-inferential doesn't mean being a gut reaction, but simply that the verdict is not inferred from some further premises. (Nor does it mean one *could* not infer it.) For parallel, we might think

---

[6] Note also that the present view is more liberal than, say, Sosa 2007, who also emphasizes understanding and competence, since he restricts competence to specifically *conceptual* competence with respect to *necessities* (though I'm open to the latter restriction).

[7] This is consistent with it being the case that some philosophers who are competent with the concept disagree about the verdict, as long as there's some alternative explanation of the disagreement (e.g. in terms of background theoretical beliefs or performance error).

[8] For the purposes of this paper, I want to remain neutral on the nature of moral competence. For some key elements of my sentimentalist account, see Kauppinen 2013 and Kauppinen 2017.

of an aesthetic judgment, which might require consider a work of art in its historical context, and viewing it together with other works to gain appreciation into what is distinctive about it. This type of reflection does not consist in drawing a conclusion from premises, and is thus consistent with one's verdict being non-inferential, even if it takes a lot of cognitive work to get to. This is why it's not misleading to talk of 'considered judgments' as intuitions.

Suppose, then, that we reflect on a case like Loop, and form a non-inferential verdict that it is permissible to kill the one person in the circumstances. Insofar as this verdict really manifests moral competence, we have some non-inferential justification to believe that the act is permissible. And this is a good reason to reject DDE, since it implies that this act is not permissible. On this picture of intuitions, it would be too strong to say that it shows DDE is false, since even competent verdicts are fallible, and the balance of reasons might still favour retaining the principle. But case intuitions still serve as an independent check on moral principles, as the standard methodology assumes. The big methodological issue that remains is just when intuitions genuinely reflect competence, especially considering the fact of persistent moral disagreement.

2. The Data: How Philosophers Think

I said that most philosophers give the same verdict about the common trolley cases. According to many studies, so do most ordinary people. It might therefore seem that there is no particular reason to doubt that, say, it is permissible to turn the trolley in Switch.

Alas, psychologists and so-called *restrictionist* or *negative* experimental philosophers have found plenty of reason to doubt it. To use the language I employed above, they hold that either people lack competence with trolley cases and other philosophical questions, or their judgments don't manifest it (they are performance errors). By now, dozens if not hundreds of studies have shown that people's verdicts about philosophically interesting cases

appear to be significantly influenced by factors other than target content, including manner of presentation and demography, which in turn suggests they can't be reliable. For example, in trolley cases, the order in which cases are presented makes a difference to verdicts (Lanteri et al. 2008), and so does the subject's mood (Valdesolo and DeSteno 2006), or variation in superficial content like racially charged names of the characters (Uhlmann et al. 2009), or the way the case is framed (Petrinovich and O'Neill 1996). Different demographic groups also seem to judge some cases differently. (I won't go into the demographic issues here, however, since unlike presentational effects, they don't directly call someone's reliability into question.[9])

One standard response – and as I will eventually argue, the correct response – is to say that philosophers are different from ordinary folk in relevant respects, more competent or better able to manifest their competence. Sometimes this is put in terms of *philosophical expertise*. As Williamson says, "the initial presumption should be that professional analytic philosophers tend to display substantially higher levels of skill in thought experimentation than laypeople do" (2011, p. 211). This line of thought is often defended by analogy with sciences – clearly laypeople's physical intuitions are likely to be worse than trained physicists'. However, the analogy argument by itself is inconclusive. Even if there are similarities, there are also pertinent differences between disciplines. As Jonathan Weinberg (et al. 2010) has emphasized, the way many experts in other areas are trained presupposes clear and immediate feedback when one makes a mistake, which makes it possible to calibrate one's intuitions. This is missing in philosophy: "philosophers' intuitions about cases do not receive anything like the kind of substantial feedback required for such virtuous

[9] For example, whether one is a liberal (in the American sense) or conservative is a demographic factor that predicts whether one believes in climate change. It doesn't, however, call into question the reliability of the liberal's belief in climate change that she would likely believe otherwise, were she a conservative.

tuning." (2010, p. 341) Even if philosophers have distinctive expertise, it doesn't follow they have better *intuitions* in particular.

I think it's fair to say that merely pointing out the analogy does not suffice. We need further reasons to believe one way or another. One potential source for such reasons is studying whether philosophers respond differently from ordinary people, and how. A few such studies have indeed been performed. Given my topic here, I'm going to focus on the three studies on trolley scenarios so far conducted on moral philosophers in particular.[10]

Two of the most relevant studies come from Eric Schwitzgebel and Fiery Cushman (2012, 2015). They presented philosophers with variants of trolley cases (and some others) varying the order of presentation. And just as with ordinary people, they found that this made a difference to philosophers' responses. Their design isn't optimal, since they're not asking whether it is morally permissible to turn the trolley or push the heavy man, but rather how morally *bad* it is to do either of these things on a scale of 1 to 7. This is unfortunate, since it is the former question that is of philosophical interest in these cases – after all, you might well think that one of two equally impermissible acts is worse than another, for example. But setting this worry aside, what S&C found is that if the Switch case was presented first, 54% rated killing one in each case as equally bad – that is, nearly half thought it is not as bad to turn the trolley as it is to push the heavy man. But when the Footbridge case was presented first, 73% rated the cases the same way – that is, they thought turning the trolley was as bad as pushing the heavy man.

---

[10] In addition to the studies I'll discuss, Tobia, Buckwalter, and Stich (2013) found that philosophers in general were vulnerable to a framing effect known as actor-observer bias in the Switch case (they were significantly more likely to think it is morally obligatory to turn the trolley if the choice was framed in first-person rather than third-person terms). However, as Joanna Demaree-Cotton (2016) has pointed out, it is not at all clear whether the difference is attributable to a 'bias', given that philosophers know morally relevant facts about themselves that they don't know of a thinly characterized third-person placeholder, so this is not a clean case of an epistemically problematic effect, and I won't focus on it here.

Schwitzgebel and Cushman (2012) also made a further potentially important finding: the order in which the cases were presented influenced also whether philosophers endorsed the Doctrine of the Double Effect – or, to be precise, a related principle I'll call DDE*, which says that intended harm is morally worse than harm caused as a foreseen side effect. Basically, philosophers, though not ordinary people, were more willing to endorse DDE* when they got the Switch case first (and thus rated causing harm as a side effect as less bad than causing it as a means) than if they got the Footbridge case first. As S&C put it, "This effect is particularly striking because, regardless of the order of presentation, all philosophers had viewed and judged the same pairs of cases by the time they were asked about the general principles." (2012, p. 149) As they see it, this suggests that rather than providing better intuitions, philosophical expertise consists in being able to *rationalize* one's intuitions better post-hoc.

In a follow-up study, Schwitzgebel and Cushman (2015) tried to nudge philosophers to reflect carefully on the scenarios to see whether this would make a difference. They gave the respondents the following prompt:

> Over the course of the five questions that follow, we are particularly interested in your reflective, considered responses. After each case, please take some time to consider the different moral dimensions at issue, including potential arguments for and against the position to which you are initially attracted. Also please consider how you might respond to different variants of the scenario or to different ways of describing the case. After you finish reading each of the five cases, there will be a 15-second delay to encourage careful reflection before you are asked a question about the case. (Schwitzgebel and Cushman 2015, p. 130)

The results were similar, though the participants in this study in both reflection and non-reflection conditions were more likely to think the cases differed. In the reflection condition, around 23% of philosophers thought turning the trolley was as bad as pushing the heavy man if Switch was presented first. If Footbridge was presented first, however, 39% thought turning the trolley was as bad. So again, the verdicts of philosophers were vulnerable to presentational effects, even though they were encouraged to engage in the kind of reflection supposedly characteristic of philosophy.

The third and most recent study I'll talk about comes from Alex Wiegmann, Joachim Horvath, and Karina Meyer (ms). Inspired by Peter Unger's response to Judith Thomson, they presented professional ethicists with the Six Options trolley case in addition to the standard Footbridge (with nine instead of five people on the straight track). The verbal description of the case is very complicated (see Unger 1996 for the somewhat simpler cases that inspired it), so I will simply reproduce their illustration:

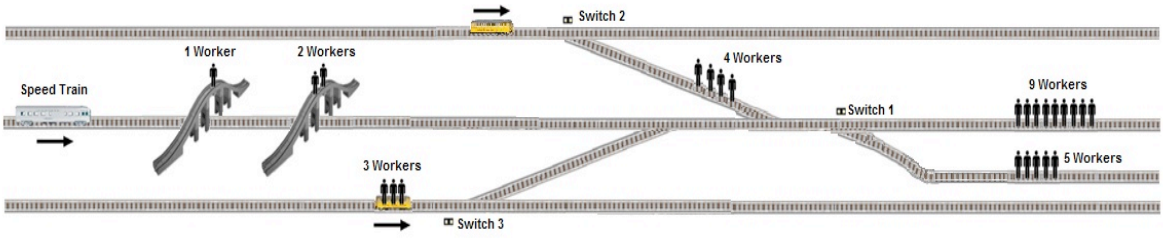

Fig.1 The Six Options Case

Here, for example, the agent can either push down one worker to stop the train, or drop two through a trapdoor, or redirect another train to stop it at the cost of killing four workers on the connecting track, or just let it kill nine. Perhaps unsurprisingly, ethicists' responses to the Six Options case were scattered. However, though only 32% thought pushing the one was acceptable in the Footbridge variant if presented first, if Six Options was presented first, 49% said pushing one was permissible in Six Options and 55% thought so in Footbridge. So

here both adding irrelevant (unchosen) options and order of presentation had a significant effect on ethicists' verdicts.

What explains these troubling effects in both ordinary people and philosophers? Perhaps the most comprehensive account is offered by Edouard Machery (2017) in his recent book on philosophical methodology. Machery emphasizes that he is not committed to radical skepticism: there's no reason to doubt people's judgment in ordinary cases. The findings don't call into question our ability to tell whether attacking civilians with chemical weapons is wrong. Instead, Machery holds that there is something about philosophical cases in particular that renders judgment unreliable. According to him, unlike situations we encounter in ordinary life, philosophical cases are beyond the *proper domain* of the relevant capacities for judgment (2017, p. 113). The first of his three main explanations for this is that in virtue of either superficial or target content, philosophical scenarios are *highly unusual*, that is, very different from the kind of cases that we have trained our judgment on. This makes it difficult to fill in the gaps in scenario descriptions, and everyday heuristics won't work. Machery suggests that uncertainty about what holds in the scenarios in part explains demographic and presentation effects, in part because the significance of superficial details is amplified in such circumstances (ibid., pp. 114-116). Second, philosophical cases "pull apart the properties that go together in everyday life" (ibid., p. 116). For example, usually, when we engage in physical violence, it does more harm than good. But in the Footbridge trolley case, violence to the one heavy man brings about the greater good of saving five. Machery observes that psychological research on categorization shows people are in general bad at categorizing atypical instances of a kind, so that can be expected to happen with philosophical scenarios as well. Finally, Machery suggests that it is difficult for people to disentangle the superficial and target content, so that their judgments are influenced also by the former (ibid., pp. 118-120).

I think these are plausible suggestions. However, I think it may be more fruitful to classify them in a different way. Let's distinguish three kinds of error:

Construal error

People who fail to grasp the target content (and thus misconstrue the case) make this kind of mistake. For example, if a subject fails to see that the heavy man is used as a means to save five in Footbridge against his consent, she makes a construal error.

Focusing error

People who do grasp the target content but whose response is influenced by superficial content in addition to (or instead of) the target content make this kind of mistake. For example, someone who thinks pushing the heavy man is wrong because there's direct physical contact, or someone who thinks it's wrong to turn the trolley because it's against the law, is guilty of a focusing error.

Classificatory error

People whose response to the target content does not manifest competence with the subject matter are likely to make classificatory errors. For example, someone who thinks that false beliefs can amount to knowledge is probably guilty of a classificatory error.

In these terms, then, the hypothesis is that people's judgment in philosophical cases is unreliable, because some features of those cases make one or the other kind of error likely enough. Perhaps the unusual or even cartoonish nature of the cases makes it hard for people to fill in implicit content, including some of the target content. Perhaps the emotional resonance or real-life associations of some cases result in focusing errors. And perhaps the fact that features that co-occur in paradigm cases are pulled apart makes it difficult to decide,

say, whether a concept applies or whether an act is permissible, resulting in classificatory errors.

We can now give a different characterization of what the relevant kind of philosophical expertise would consist in: having expertise with thought experiments is being disposed to avoid construal, focusing, and classificatory errors. On the face of it, the training that philosophers standardly receive can be expected to give rise to such expertise. When it comes to construal, philosophers read, discuss, and construct many scenarios in the course of their studies, including variations of the same type of case. Many acquire the ability to figure out which details are superficial for the philosophical purpose at hand, and to zoom in on the pertinent features. What is unusual for non-philosophers – such as complex entities created in an instant or entire universes containing only one object – is humdrum for philosophers. When it comes to focusing, philosophers are taught from the first to sharply distinguish the different sort of questions that might be asked about a subject matter, say legal and moral ones, and correspondingly to focus on what might be relevant to answering them.

Classification is trickier in the absence of uncontroversial feedback for getting it wrong. No liquid is going to change colour if you misclassify a possible action as permissible. (This is a persistent problem for accounts that treat moral intuitions on the model of empirical intuitions, such as Railton 2014.) But at least philosophers have well-established distinctions at their disposal to help them make fine-grained classifications, whether it is between duty and supererogation, violating and infringing rights, or prima facie and pro tanto reasons. For example, it might be that both a layperson and an ethicist have an ambivalent feeling about Switch, because both tacitly recognize that while the greater good would be served by turning the trolley, the one on the side track has a right not to be killed. But the ethicist, being trained to distinguish between permissibly infringing a right and impermissibly violating one, might recognize the scenario as a case of the former. At least

equally importantly, thinking about the point of a classification can be expected to help make it. We don't have the concepts we do by chance, but because we need them to do a job. And what qualifies as an X may depend on what follows from being an X. For example, if doing something is permissible, but there is nevertheless a strong pro tanto moral reason against doing it (say, breaking a promise to achieve some important good), there will typically be residual obligation to make amendments. Recognizing this will help classify something as a pro tanto reason. In Machery's terms, the training philosophers receive can be presumed to expand the proper domain of their relevant capacity for judgment.

As I'll briefly argue below, in addition to what I've said about the effects of individual training, the chances of all three types of error may well be reduced by the way philosophy functions as social practice. The question, then, is whether the empirical data regarding philosophers undermines the presumption that philosophers, individually or as a community, have expertise of the right kind with thought experiments.

3. Facing the Challenge

So, given that none of the empirical studies performed on philosophers show any sign of their intuitions being better – or even different – from the shaky intuitions of ordinary people, should we not stop appealing to their content, and relinquish the parts of philosophical practice that require doing so? One influential response to this concern has been arguing that it relies on a false assumption: intuitions are not a load-bearing part of philosophical practice in the first place. Thus, Herman Cappelen (2012) argues that philosophers don't in their actual practice rely on propositional attitudes that have a special phenomenology or are based solely on conceptual competence, or on non-inferential and non-experiential judgments that are recalcitrant to evidence and serve as rock bottom justifiers, which suggests they don't appeal to 'intuition', in spite of appearances (cf. Deutsch 2015). For

example, in discussing Thomson's (1985) use of the Switch case, Cappelen denies that she treats it "as a point where justification gives out", and claims that instead, the "goal of the paper is to look for reasons and evidence beyond the pre-theoretic judgment" (2012, p. 161). As we'll see, I think both these points are correct. Nevertheless, it doesn't follow that contents of intuitions play no important evidential role. Even if Thomson and others offer non-intuitive premises which entail their verdicts about cases (which they do), I believe it is clear that at least part of the justification for the non-intuitive premises is meant to be that they fit with the intuitive verdicts. For example, part of her case for the claim that it makes a difference to permissibility whether the agent needs use means that themselves infringe someone's rights in order to save a larger number of people is that it (purportedly) explains the difference between Switch and Footbridge, which she clearly treats as a tentative given independently of her particular explanation.[11]

However, while I think intuition deniers go too far, their work points to something crucial about philosophical practice: thought experiments are never performed outside a theoretical context. Nor is it true, generally speaking, that intuited propositions are treated as having rock bottom epistemic status, even if they have some independent standing. Properly appreciating both of these aspects of the use of intuitions will be essential to responding to the restrictionist challenge, or so I'll argue.

3.1 Thought Experiments In and Out of Context

I'll begin with the role of theoretical context in the actual practice of appealing to intuitions. What I'll argue is that it raises serious doubts about the *ecological validity* of the studies – that is, whether we have good reason to think that results about the subjects' behavior in the experimental condition generalize to real-world conditions. After all, it should be

[11] For a particularly careful argument to this effect, see Nado 2017.

uncontroversial that expertise only gets exercised in certain conditions. Being an expert at something doesn't mean you'll always perform at full tilt. So one potential, and in my mind quite likely, debunking explanation of the studies is that in the experimental condition, the experts don't, or rather can't, make use of their expertise.[12] Just think of how different from actually employing the method of cases the experimental condition is. There you are, at your computer, and you're given a scenario, and then another. You're not given a theoretical context. When you encounter a scenario, you're not engaged in contemplating whether or when it is right to harm some to help others, or why that is, for example. It's not clear what hangs on your answer. You might want to reflect for a bit (at least 15 seconds) to be helpful, but without knowing the big picture, it's difficult if not impossible to do.[13] Philosophers no more conduct thought experiments without a particular hypothesis in the back of their mind than chemists mix up substances in the lab without some idea of what might happen.[14]

Consider Switch in this light. As I observed in the first section, the context in which Thomson (1985) introduces it is the non-utilitarian family dispute on whether it is always impermissible to do harm, while it is permissible to allow it to happen. It is tempting for a nonconsequentialist to think so, since such a principle makes sense of many everyday cases.

---

[12] Rini (2015) observes along the same lines that "Proponents of the expertise defense are certainly not committed to the view that trained philosophers have an ability to always, everywhere, under any circumstances, avoid intuition-distorting effects" (p. 444). She uses this to explain the curious fact that ethicists' responses to trolley cases vary, even though they should be very familiar to them (and thus pose a different task for them than for ordinary subjects).

[13] This is also my hypothesis why simply being induced to reflect in various ways doesn't have a significant effect on ordinary people's judgments, as Colaço et al. (ms) found.

[14] A secondary consideration concerns the practical context. Nothing is at stake for you personally when you're answering an anonymous survey of professional philosophers. You might have some level of curiosity and goodwill, but if you're at all like me, you'll try to complete the task as fast as possible and move on to Netflix, in part because in the absence of the theoretical context, you don't really know what the point of the thought experiment is, and you don't really know what to say. If you were presenting a case or publicly responding to it in the course of a discussion, your professional reputation would be on the line – you wouldn't want to be seen to lack acuity and competence. But that's not what's happening here.

That's the idea the thought experiment is meant to help you evaluate. And the response Thomson means you to have, bearing in mind the tempting principle and the cases that it would explain, is that it is nevertheless permissible to *do* harm in a case like Switch rather than *allow* more harm to take place. My hypothesis is that it is in this sort of context of understanding the question, grasping the motivations for the competing views, and bearing in mind other cases that are in some respects similar that philosophers can be expected to manifest their competence with thought experimentation. To be sure, as Regina Rini (personal communication) pointed out to me, this also means that the conditions in which philosophers are best positioned to conduct thought experiments will often be the very ones in which they are most likely to be biased in virtue of an antecedent commitment. But that is the fine line we have to walk, and the difficulty of doing so may have a major role of explaining why philosophical disagreements are so persistent.

Given the above, I'm inclined to say that the experimental condition is just about as different from serious philosophical use of thought experiments as the trolley cases themselves are from everyday moral choices. And if this is right, the results of the studies may not generalize to the actual practice of using thought experiments, even if they hold for the experimental situation. Now, some might reply: "How, then, can we test for expertise if not studies like these?" To be honest, I'm not particularly convinced that there *is* any way of testing for expertise with cases, except making use of your own judgment regarding the judgments of others. If that's no good, maybe peer review! (I personally believe that Jeff McMahan and Frances Kamm, for example, have expertise with thought experiments on the basis of having read many of their books – while my evidence is *a posteriori*, it hardly meets the criteria of a psychological experiment.)

3.2 The Salience Hypothesis of Intuitive Variation

My second line of response is an empirical hypothesis about what actually explains the observed variation. If the hypothesis about which features of the experimental situation play a crucial role is correct, and the influence of those features is absent or mitigated in ordinary philosophical practice, this should further increase our confidence in thought experiments.

I'll contrast my hypothesis with that of Schwitzgebel and Cushman (henceforth, S&C) (2012). Here's what they say regarding order effects in particular:

> We suggest that order effects arise from an interaction between intuitive judgment and subsequent explicit reasoning: The intuition elicited by the first case becomes the basis for imposed consistency in the second case […] When the intuition elicited by one case is 'stronger'—that is, more resistant to revision by explicit reasoning—than the intuition elicited by the complementary scenario, this would lead to the asymmetric equivalency effects that we report here. (S&C 2012, p. 148)

This is a relatively uncharitable explanation, since it appeals to a kind of rationalization – "imposed consistency" – in response to a recalcitrant intuition. But I think an alternative explanation is more likely, and gives reason for some optimism. It is that when many morally relevant factors are present, varying the *salience* of a factor weighing in one direction makes a difference to weak intuitions in particular. Part of the story will be the assumption that some intuitions are stronger than others – in particular, in trolley cases, the intuition that pushing down the heavy person in Footbridge is wrong is stronger than the intuition that it is okay to turn the trolley in Switch (see Zamzov and Nichols 2009 for empirical evidence).

So, here are some of the most obvious morally relevant factors present in the standard trolley cases:

- Pro tanto moral reason to benefit more rather the fewer people.

- Pro tanto moral reason not to harm others against their will.
- Pro tanto moral reason not to use others as means to help others.
- Pro tanto moral reason not to interfere with the physical integrity of others.

As Machery's explanation of folk responses says, these reasons usually point in the same direction – usually, the reason not to harm others favours the same act as the reason to benefit more rather than fewer people. But in trolley cases they point in different directions, forcing us to weigh them against each other in arriving at a permissibility judgment. Especially when our response is non-inferential, as it is by definition when it is a case intuition, this weighing can be expected to be sensitive to how salient each kind of reason is, and not just the strength of the reason. This salience hypothesis predicts the observed effects. Here's how. Call the sequence in which Footbridge precedes Switch 'Footbridge First'. In Footbridge First, the reason not to harm others against their will is highly salient, and associated with reasons not to use others as means or interfere with them physically. Since Switch also involves harming others against their will, raising the salience of this reason can be expected to get people to regard turning the switch more negatively, if the permissibility intuition is weak. It is, so unsurprisingly, quite a few people have the intuition that turning the trolley is impermissible. In contrast, in a Switch First sequence, the reason to benefit more people even if fewer people are harmed is highly salient. Since Footbridge also involves benefiting more rather than fewer people, raising the salience of this factor can be expected to get people to regard pushing the man more positively, if the impermissibility intuition is weak. However, that's not the case: the impermissibility of pushing is a robust intuition. And indeed, the effect is not found.

The next part of the defensive case is that appealing to variation in salience as an explanation of order effects – or framing effects for that matter – is a relatively *benign*

explanation of unreliability, because it is something that can be (and is) counteracted. What's going wrong is that raising salience causes improper weighing of the different reasons in intuitive verdicts. But the intuitions are nonetheless responsive to genuine reasons, which would arguably be sufficient reasons, if they were the only relevant features. This kind of mistake can be avoided in principle and in careful practice. For example, it's a good idea to consider the Switch case side by side not only with Footbridge, but also, say, rescue cases, in which one has to choose between saving one or many people. Perhaps there are yet other relevant features that are highlighted by considering other types of case. That's precisely how nonconsequentialists like Frances Kamm (2015) and Judith Thomson (1985; 2008) actually proceed in their work. It doesn't make for great prose, but it seems to be an effective way to guard against being misled by variable salience.

Notice also that since philosophical training involves considering a variety of different scenarios, standard models of inductive or statistical learning suggest that philosophers should be capable of picking up patterns along the lines of 'using someone as a mere means is a wrong-making feature', even if each individual case is 'noisy' in the sense that many different features are potential (or even actual) contributory factors to accounting for the overall intuitive verdict. To put this differently, even if someone like Machery is right in saying that in the properties that philosophical thought experiments isolate, say A, B, and C, go together in everyday cases, a varied diet of everyday cases in which these and other properties form different combinations can teach a careful observer that A by itself makes some difference.

Considering the Six Options case suggests a different possible remedy that is also part of standard practice. Given general constraints on human cognition, if we want to find out which potentially relevant features make a moral difference, we need to focus on pairs of scenarios that differ only along one dimension. Especially outside the context of a specific

theoretical question, very few if any of us are capable of responding non-inferentially to simultaneous comparison of six different options, not to mention comparing each option with every other pairwise. Speaking for myself, the only way I can figure out what is permissible in Six Options is explicit reasoning in light of principles supported by pairwise comparison of simpler cases – and even then, I'm not confident in my judgment. (For the record, the correct option seems to be hitting switch 3, redirecting the train carrying three workers to the main track – assuming that the weight of the small train suffices to stop the speed train by itself, so that the three workers die as a side effect of stopping the threat to nine. See e.g. Kamm 2015 for why this would be the right thing to do.)

What explains the relative popularity of willingness to push down the one person in Six Options is most likely that it is clearly better than one of the other options along one relevant dimension, namely the option of dropping down two people, as well as better than all the rest along the dimension of saving lives. In psychological terms, the option to drop two people is a decoy option – it invites us to make an easy comparison that makes another option, in this case pushing the one, to look good, when we find it otherwise difficult to collapse all the different dimensions into one ordinal ranking (see Ariely 2008, pp. 1–10). Such additional options have been found to influence judgment in trolley cases – as Shallow, Iliev, and Medin (2011, p. 598) put it, "Adding a third option to a binary choice set selectively interferes with the approval rating of the closest alternative, a pattern consistent with similarity effect." It is thus not surprising that philosophers are tricked by Six Options, since their cognitive limitations make it practically impossible to form a reliable non-inferential verdict. The only remedy to this is refusing to even try to form intuitions about such scenarios – which is just what most ethicists do as a matter of fact (in effect, they rightly refuse to play Unger's game).

The two replies so far together suggest the following hypothesis:

Hypothesis

The more closely the experimental condition resembles ordinary philosophical reflection, the less difference there is in the salience of the different target features of the scenario, and the stronger the intuition is, the less philosophers' case intuitions are influenced by presentational factors.

If Hypothesis is true, current empirical data regarding philosophers' intuitions says very little about the reliability of the non-inferential verdicts made in the context of actual philosophical practice.

3.3 Screening Intuitions

So far, I've argued that we have good reasons to think that the striking fallibility of philosophers' verdicts in experimental situations doesn't generalize to the actual use of thought experiments. But what if it does? At this point, the coherentist nature of philosophical methodology comes to play an important role. Intuition deniers are right in denying that philosophers accept intuitions at face value, or treat them as justifiers whose status is beyond question. Rather, it is standard practice to try to fit intuitions about cases with general principles that have an independent rationale (see e.g. Kagan 1989 and Kamm 2015). Both of these moves, I'll argue, are likely to mitigate the remaining effect of irrelevant factors. As Regina Rini (2015, p. 434) puts it, "Perhaps philosophers are no more likely than anyone else to have good intuitions, but are much better at judiciously using the ones they do have."

One part of the use of intuitions that is worth noting to begin with is that philosophers are often aware of the relative strength of different intuitions, and take this into account in the weight they give to them in theory construction. Consider what Jeff McMahan

says regarding his Conscientious Driver case. Briefly, he holds that it is morally permissible for an innocent pedestrian to kill a careful and conscientious driver in self-defense, when the driver would otherwise kill her in a freak accident. This fits with McMahan's Responsibility Account of liability, according to which, roughly, agents are morally liable to harm when they are responsible for choosing to engage in an activity that foreseeably imposes a risk on others. After considering variations of the case, he concludes:

> I continue to believe … that the Responsibility Account's implications for the case of the conscientious driver are more plausible than those of the Culpability Account [which entails that the driver is not liable to harm, since he's not culpable – AK]. I do not, however, point to this in triumph, for this is not the kind of case against which we can usefully test a theory's implications. It is a case about which most of us have only weak or doubtful intuitions, and thus is precisely the sort of case for which we need guidance from a theory. (McMahan 2005, p. 403)

Here McMahan treats his thought experiment with appropriate caution, giving it little weight on its own. (And indeed, many critics flat out reject his verdict on the case.) One reason this matters is that empirical research has shown that weak intuitions are most vulnerable to situational effects. In philosophical context, Jennifer Wright (2010) found that in epistemological cases, the more confident ordinary people were of their judgments, the less likely it was that they were subject to situational variation. And Zamzov and Nichols (2009) found the same for trolley cases. So insofar as caution in proceeding with weak intuitions is already a part of philosophical practice, it is likely to mitigate the effects of unreliable intuitions.

Nevertheless, there is no guarantee that caution suffices to screen out all shaky intuitions. If ethicists had nothing but case intuitions to go on, they would have to gamble on

the absence of performance error without any independent criteria for assessing whether this is the case. That does not, however, correspond to their actual practice. General ideas like 'it's typically morally good to make choices that make as many people as possible better off' or 'it's typically morally bad to treat people as mere tools in the pursuit of your own advantage' are both plausible on their own and apt generalizations from everyday cases (indeed, the latter may explain the former). Such ideas serve as building blocks for candidate moral principles that should not be lightly discarded and can serve as independent checks for case intuitions. Again, this is not to say they have rock bottom status either. It suffices that neither candidate principles nor case intuitions derive their justification solely from coherence with each other, but each comes to the balancing process with independent credibility. Non-inferential and inferential justification can be mutually supportive. As Mark van Roojen (2014) has argued, even if the initial credibility of both case intuitions and candidate principles is insufficient to justify belief in either, their fit with each other may increase the justification of each to the level at which belief is warranted.

Famously, some argue more strongly that basic moral principles are self-evident (Sidgwick 1907, Audi 2015) or can be derived from practical rationality or the conditions of agency (Kant 1996, Korsgaard 2009). Whether or not these ambitious research programmes are successful, candidate principles that are not just generalizations from intuitions about fanciful cases are not subject to experimentally motivated criticism, so the odds of a garbage-in, garbage-out reflective equilibrium process are reduced.

In this respect, it is notable that Schwitzgebel and Cushman's original 2012 finding of post hoc rationalization did *not* replicate in their follow-up study (2015, p. 133). That is, while they still found order effects in philosopher's verdicts about cases, this did not make a statistically significant difference to whether they endorsed the variant of DDE* they tested.

This is cause for optimism, as it suggests that even in imperfect experimental conditions, the principles philosophers endorse aren't simply based on case intuitions.

3.4 Intuition Processing as a Social Practice

If we participate in a good enough epistemic practice, we may together know what I wouldn't know by myself, so that *I* may know something because *we* know it. This is how I know that the Earth revolves around the Sun, for example, since science and science-based education are good enough epistemic practices for me to know such things. So the question is: is philosophy a good enough epistemic practice to allow me to know that it's wrong to push the large man off the footbridge, for example? There is, of course, a lot that is required for an epistemic practice to be good, and such requirements will plausibly vary with the subject matter (see, in general, List and Pettit 2011). There is much work to be done on the social epistemology of philosophy. Here, I'll just focus on a few pertinent aspects, mutual correction and deference.

When it comes to the use of thought experiments, it might be that we initially respond to cases alone, whether as producers or consumers. But these responses are then put to use to construct theories or convince others. And given the way the discipline is structured, those others are not going to take it at face value that things are as you say. It is a truth universally (if sometimes ruefully) acknowledged that analytic philosophy is adversarial. If I fail to construe or focus on or classify the target content of a case properly, the odds are very high that some smart person will call me out. That's not always nice, of course, but it's conducive to getting it right. After all, it doesn't matter so much if *I* hold false beliefs about what's permissible and why. It's much more important what *we* as a profession think and

teach our students. And a practice in which only the strong intuitions survive is conducive to avoiding mistakes arising from circumstantial variation, at least.[15]

What's more, as in other epistemic practices, some practitioners are widely acknowledged to be specialists in thought experimentation, and treated with some measure of deference. Schwitzgebel and Cushman explicitly say that they don't think that the intuitions of super-specialists like Thomson or John Martin Fischer are subject to the kind of variation they find (2012, p. 136). Insofar as lesser lights such as myself treat the case intuitions of such thinkers as having a high initial credibility, this again reduces the chances of a mistake in the context of actual practice (but not in the experimental context).

So suppose I think it's wrong to turn the trolley in Switch, because I encountered Footbridge first. It's highly unlikely that the same goes for everyone. In fact, we know most don't have this intuition. Is the best explanation for this that most people encountered Switch first? Again, that's highly unlikely. Or suppose that I'm led by the superficial difference between pushing a heavy man and opening a trap door by hitting a switch to give a different verdict to Footbridge and Trapdoor. In the absence of some further rationale, this is not going get past a murder of philosophers.[16] People will push you for a relevant difference, and they won't easily accept something that fails to carry weight in other cases with a similar causal structure. And they certainly won't defer to you. So whether it comes to construal, focusing, or classification, our verdicts about cases are going to face pushback in the context of ordinary philosophical practice, and that should increase our confidence in surviving propositions, even independently of whether they have further argumentative support.

---

[15] Of course, there is no guarantee this always works, and other factors like prestige and lack of demographic diversity can skew which verdicts become established data points, but at least there's some positive tendency here.

[16] The collective noun for a group of crows is 'a murder'; I'm here suggesting that the same term can be applied to a group of philosophers, for obvious reasons.

4. Conclusion: The Presumption Survives

When I was a graduate student, I had the privilege of taking a semester-long seminar with Alfred Mele at FSU. In the course of discussing action theory, moral responsibility, and free will, Al took us through dozens of thought experiments, pushing us to try out variations and come up with principles. And whenever I was foolish enough to propose a principle, he immediately presented a devastating (and often funny) counterexample, and modified it if I modified my principle to capture the original variant. To an extent, you might say it was a kind of a game, but it had a serious purpose: it forced me (and other students) to think through the implications of commitments, to imagine systematically and precisely how things might be, to ignore the accidental, and to embrace a degree of intellectual humility. And while Al's seminar was a master class, there's no reason to think it was entirely exceptional.

Critics of standard methodology in normative ethics and elsewhere in philosophy, dismiss the effectiveness of such training[17] in part on a posteriori grounds, on the basis of studies that appear to show philosophers are no less vulnerable to presentational effects than laypeople. I've argued that the lesson of these studies is different. It is that it would be a mistake to think that philosophical training fosters a skill in thought experimentation that can be exercised in isolation from the context of a philosophical debate, just by looking at and reflecting on a scenario. But that is an assumption that should never have been made, part of a mythical conception of 'the method of cases'. Ethicists and philosophers in general use thought experiments in the context of articulating, defending, and attacking specific hypotheses, so that the kind of expertise they have with intuitions isn't independent of the

[17] Indeed, Machery (2017, p. 164) even denies it exists!

other kinds of expertise that even intuition sceptics acknowledge they have.[18] In consequence, competent intuitive verdicts about ordinary and unusual scenarios offer some support to purported moral principles, but are nevertheless subject to critical scrutiny both at the individual and social level. So while questions about persistent disagreement and cultural variation remain, as far as empirical studies of ethicists go, there's no need for philosophers to be afraid of trolleys, trains, or ruthless surgeons – at least as long as they're only imaginary.[19]

References

Ariely, Dan (2008) *Predictably Irrational* (New York: HarperCollins).

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

Bourget and Chalmers (2014) 'What Do Philosophers Believe?' *Philosophical Studies* 170 (3), pp. 465–500.

Cappelen, Herman (2012) *Philosophy Without Intuitions* (Oxford: Oxford University Press).

Colaço, David, Kneer, Markus, Alexander, Joshua, and Machery, Edouard (ms) 'On Second Thought: A Refutation of the Reflection Defense'.

Dancy, Jonathan (1993) *Moral Reasons* (Oxford: Blackwell).

Demaree-Cotton, Joanna (2016) 'Do Framing Effects Make Moral Intuitions Unreliable?' Philosophical Psychology 29 (1), pp. 1–22.

[18] Timothy Williamson (2011, p. 222) makes a similar point in observing that conducting a thought experiment is a special case of argument construction and evaluation, which even skeptics like Weinberg et al. (2010) acknowledge is part of philosophical training.

[19] I'd like to thank Joachim Horvath, Edouard Machery, Jennifer Nado, Lilian O'Brien, Stephen Stich, Jussi Suikkanen, Alex Wiegmann, and participants at workshops in Birmingham, Warwick, Bielefeld, Dublin, and Helsinki for comments and criticisms. Special thanks go to Regina Rini for agreeing to 'peer review' my manuscript for this volume, and for her very thoughtful written comments.

Deutsch, Max (2015) *The Myth of the Intuitive* (Cambridge, MA: MIT Press).

Foot, Philippa (1967) 'The Problem of Abortion and the Doctrine of the Double Effect', *Oxford Review* 5, pp. 5–15.

Jaworska, Agnieszka and Tannenbaum, Julie (2018) 'The Grounds of Moral Status', The Stanford Encyclopedia of Philosophy (Spring 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2018/entries/grounds-moral-status/>.

Kagan, Shelly (1989) *The Limits of Morality* (Oxford: Oxford University Press).

Kahane, Guy (2012) 'On the Wrong Track: Process and Content in Moral Psychology', *Mind and Language* 27 (5), pp. 519–45.

Kamm, Frances (2015) *The Trolley Problem Mysteries* (New York: Oxford University Press).

Kant, Immanuel (1996), *Practical Philosophy*, Mary Gregor (ed.) (Cambridge: Cambridge University Press).

Kauppinen, Antti (2007) 'The Rise and Fall of Experimental Philosophy', *Philosophical Explorations* 10 (2), pp. 95–118.

Kauppinen, Antti (2013) 'A Humean Theory of Moral Intuition', *Canadian Journal of Philosophy* 43 (3), 360–81.

Kauppinen, Antti (2017) 'Sentimentalism, Blameworthiness, and Wrongdoing', in *Ethical Sentimentalism,* Karsten Stueber and Remy Debes (eds.) (Cambridge: Cambridge University Press), pp. 133–52.

Korsgaard, Christine (2009) *Self-Constitution* (New York: Oxford University Press).

Lanteri, Alessandro, Rizzello, Salvatore Marco, and Chelini, Chiara (2008), 'An Experimental Investigation of Emotions and Reasoning in the Trolley Problem', *Journal of Business Ethics* 83 (4), pp. 789–804.

List, Christian and Pettit, Philip (2011) *Group Agency* (Oxford: Oxford University Press).

Machery, Edouard (2017) *Philosophy Within Its Proper Bounds* (New York: Oxford University Press).

McMahan, Jeff (2005), 'The Basis of Moral Liability to Defensive Killing', *Philosophical Issues* 15, pp. 386–405.

Mercier, Hugo and Sperber, Dan (2017) *The Enigma of Reason* (Cambridge, MA: Harvard University Press).

Moore, George Edward (1903) *Principia Ethica* (Cambridge: Cambridge University Press).

Nado, Jennifer (2017) 'Demythologizing Intuition', *Inquiry* 60 (4), pp. 386–402.

Petrinovich, Lewis and O'Neill, Patricia (1996), 'Influence of Wording and Framing Effects on Moral Intuitions', *Evolution and Human Behavior* 17 (3), pp. 145–71.

Railton, Peter (2014) 'The Affective Dog and Its Rational Tale' *Ethics* 124 (4), pp. 813–59.

Rini, Regina (2015), 'How Not to Test for Philosophical Expertise' *Synthese* 192(2), pp. 431–52.

van Roojen, Mark (2014), 'Moral Intuitionism, Experiments, and Skeptical Arguments', in *Intuitions*, Anthony Robert Booth and Darrell P. Rowbottom (eds.) (Oxford: Oxford University Press), pp. 148–64.

Schwitzgebel, Eric and Cushman, Fiery (2012) 'Expertise in Moral Reasoning? Order Effects on Moral Judgment in Professional Philosophers and Non-Philosophers', *Mind and Language* 27 (2), pp. 135–53.

Schwitzgebel, Eric and Cushman, Fiery (2015) 'Philosophers' Biased Judgments Persist Despite Training, Expertise and Reflection', *Cognition* 141, pp. 127–137.

Shallow, Christopher, Rumen Iliev, and Douglas Medin (2011), 'Trolley Problems in Context', *Judgment and Decision Making* 6 (7), pp. 593–601.

Sidgwick, Henry (1907) *The Methods of Ethics* (London: Macmillan).

Sosa, Ernest (2007) *A Virtue Epistemology* (New York: Oxford University Press).

Thomson, Judith (1985) 'The Trolley Problem', *Yale Law Journal* 94, pp. 1395–1415.

Thomson, Judith (2008) 'Turning the Trolley', Philosophy and Public Affairs 36 (4), pp. 359–74.

Tobia, Kevin, Buckwalter, Wesley, and Stich, Stephen (2013) 'Moral Intuitions: Are Philosophers Experts?' *Philosophical Psychology* 26, pp. 629–38.

Uhlmann, Eric Luis, Pizarro, David A., David Tannenbaum, and Peter H. Ditto (2009), 'The Motivated Use of Moral Principles', *Judgment and Decision-Making* 4 (6), pp. 476–91.

Unger, Peter (1996) *Living High and Letting Die* (New York: Oxford University Press).

Valdesolo, Piercarlo and DeSteno, David (2006) 'Manipulations of Emotional Context Shape Moral Judgment', *Psychological Science* 17, 476–7.

Weinberg, Jonathan, Gonnerman, Chad, Buckner, Cameron, and Alexander, Joshua (2010), 'Are Philosophers Expert Intuiters?' *Philosophical Psychology* 23 (3), pp. 331–55.

Wiegmann, Alex, Horvath, Joachim, and Meyer, Karina (ms), 'Intuitive Expertise and Irrelevant Options'.

Williamson, Timothy (2007) *The Philosophy of Philosophy* (Oxford: Blackwell).

Williamson, Timothy (2011) 'Philosophical Expertise and the Burden of Proof', *Metaphilosophy* 42 (3), pp. 215–29.

Zamzov, Jennifer and Nichols, Shaun (2009) 'Variations in Ethical Intuitions', *Philosophical Issues* 19 (1), pp. 368–88.