Mark Schroeder
University of Southern California
October 18th, 2019

A Common Subject for Ethics

The purpose of this paper is to conceptualize and explore what I shall call the Common Subject Problem for ethics. The problem is that there seems to be no good answer to what property everyone who makes moral claims could be talking and thinking about. The Common Subject Problem is not a new problem; on the contrary, I will argue that it is one of the central animating concerns in the history of both metaethics and normative theory. But despite its importance, the Common Subject Problem is essentially invisible on many contemporary ways of carving up the problems of metaethics and normative ethical theory.

My aim, therefore, is to make progress – in part by naming the problem, but also by beginning to sketch out the contours of what gives the problem its force, by distinguishing between different paths of response to the problem and assessing some of their chief merits, and finally, by distinguishing the Common Subject Problem from another problem with which it has come to be conflated. This nearby problem is the Moral Twin Earth Problem. Whereas the Common Subject Problem is a problem about what property 'wrong' could refer to, the Moral Twin Earth Problem is a problem about how 'wrong' could refer to it.

The upshot of the paper, therefore, is to rescue one of the historically significant problems in normative ethics and metaethics – a problem that is essentially about normative semantics – from the illusion that has persisted over the last twenty years that it is really, somehow, a problem about metasemantics. Once we have reclaimed this problem, we can see that it could still be a problem even if there are no distinctively metasemantic problems in metaethics at all, that it is a problem faced by a wider variety of views, and that the space of possible solutions is much wider and more interesting for normative theory, moral psychology, and moral epistemology.

I   First Pass

The Common Subject Problem, at its core, is a problem for any view that claims that there must be some substantive answer to what the common referent is of talk and thought about what is right and wrong. The problem is to identify what this common referent could be, so that people who think that abortion is wrong

all think that abortion has this property, and people who think that abortion is not wrong all think that abortion lacks this property.

Of course, there is one obvious and true answer as to what property is at stake in all talk and thought about what is wrong – it is the property of *wrongness*. And that answer is correct, so far as it goes. But only a primitivist about wrongness – someone who claims that wrongness is a simple, irreducible property in the way that Moore claims of intrinsic goodness – can rest content with this answer. But not even Moore was a primitivist about wrongness. Moore thought that wrongness was the property of failing to maximize the good. The Common Subject Problem is faced by anyone else who thinks, like Moore, that there is something substantive that can be said about what the property of wrongness is.[1]

Obviously not all theorists believe that talk and thought about what is right and wrong has a common referent in this way. Contextualists allow that people really do talk past one another, and relativists and expressivists will say that moral thought and talk has a common subject matter only in some more relaxed (but still, by their lights, respectable) sense. So such views do not face the Common Subject Problem. Indeed, I suggest, the Common Subject Problem is plausibly one of the most important and underappreciated motivations for such views, since primitivism about wrongness is not widely viewed as an acceptable alternative.

Although the Common Subject Problem is supposed to be a general problem for *any* proposed referent of 'wrong', the best way to get an intuitive feel for the problem is to start with simple and familiar possible answers to what this property might be, and what these answers seem unsatisfactory. So consider the case of constitutive hedonistic utilitarianism, a relaxation of Bentham's analytic hedonistic utilitarianism. According to this view, as according to Bentham, *what it is* to be wrong is to fail to maximize the long-run balance of pleasures over pains, and what is right is its contradictory. But in contrast to Bentham, this is not an analytic definition of 'right' or 'wrong' that is within the grasp of competent speakers – it is just the property in the world that speakers happen to have latched onto, knowingly or not, in using these words.

This makes constitutive hedonistic utilitarianism a paradigm – and the most familiar paradigm – of the sorts of view that face the Common Subject Problem. It identifies the property of wrongness by telling us, in other terms, what it is to be wrong. Constitutive Hedonistic Utilitarianism also happens to be a form of what is known as *reductive realism* in metaethics, according to which normative properties can be fully

---

[1] In contrast to the Common Subject Problem, the Moral Twin Earth Problem is claimed by its proponents to be a problem only for synthetic naturalist views in metaethics, and not for non-reductive views like Moore's.

analyzed in non-normative terms.[2] But as I have explained, this is not essential for the problem, which could just as well be illustrated with a non-analytic form of Moore's constitutive consequentialism.

Constitutive hedonistic utilitarianism does offer at least a *prima facie* plausible account of what Jeremy Bentham and John Stuart Mill are talking and thinking about, when they assert and entertain claims about what is right or wrong. This is not only because both explicitly tell us that this is what they are talking about — it also makes excellent sense of their practice. When Bentham and Mill are confronted with a moral problem, they turn their attention toward consequences. And the consequences about which they worry concern how the action affects the pleasures and pains experienced by individuals.

In contrast — and this is the problem — constitutive hedonistic utilitarianism is much less intuitively plausible as an account of what Immanuel Kant or Judith Jarvis Thomson are talking or thinking about, when they assert or entertain claims about what is right or wrong. The problem with Kant and Thomson is not simply that they have *false* views about the referents of the terms 'right' and 'wrong', which of course is true of anyone who accepts a false analysis, so much as that their views are so misguided, if constitutive hedonistic utilitarianism is true, that it is unclear how any of their ways of forming beliefs about what is right or wrong make any sense at all.

Much more remains to be said, before it is clear where the problem lies, or even if there is a true problem here at all. But this is the right level of depth to get the dialectic surrounding the Common Subject Problem going. My goal so far is just to gesture toward an intuitive judgment — that constitutive hedonistic utilitarianism somehow *seems terrible* as an account of what Kant and Thomson are talking and thinking about.

It will be helpful to distinguish between three main responses to the problem. *Dividers* tend to find it obvious that constitutive hedonistic utilitarianism gets Kant's and Thomson's thoughts deeply wrong, and to assume that this thought will generalize to any other constitutive view of the nature of wrongness. Dividers conclude by *dividing*: that there is no single property at stake in all thought and talk about what is wrong.

Uniters, in contrast, hold that the Common Subject Problem is a real problem that imposes a constraint on the answer to which property is picked out by 'wrong', because some candidates for the referent of 'wrong' are better than others simply in virtue of making better sense of the range of reasonable views, including views like Bentham's and Thomson's. Although Uniters can disagree about which are the better candidates, it is typical of Uniters to think that insofar as you have the intuitive reaction that constitutive

---

[2] The best way to define reductive realism in metaethics is contentious; here for simplicity I follow Schroeder 2007, chapter 4. Nothing in this paper turns on how we define this term, because it is my thesis that the Common Subject Problem does *not* turn on it.

hedonistic utilitarianism is an implausible account of what Thomson is thinking about, that is because this is one of the views that does poorly by this metric.

Finally, *Deflaters* tend not to feel the force of the problem – or to feel like whatever intuitive force it has, there can't be a real problem, here, because insofar as they are able to articulate what the problem is in theoretical terms, it seems to have straightforward answers.[3] If you are thinking right now that I am belaboring the obvious, you are probably a Divider. If you are thinking right now that I have failed so far to describe any real problem, then you are probably a Deflater. Either way, my goal in this paper is to encourage you to take Uniting more seriously.

2   Second Pass

So far, I've given a superficial, first pass at the Common Subject Problem. But Deflaters tend not to see the force of the problem, so it merits sharpening substantially. I do *not* aim to argue against Deflating as such in this paper, but I do want to argue that even Deflaters should accept that there is at least a *prima facie* puzzle, which merits a solution. It is to that task that I turn, in this section, in giving a second, deeper, pass at the puzzle. My goal here is therefore not to argue that there is *no* answer that the constitutive hedonistic utilitarian can give to the problem – indeed, in section 3 I will go on to consider some of the more promising modes of response, some of which are available to constitutive hedonistic utilitarianism. Rather, my goal is simply to show that any answer to the problem will be non-trivial – that it is at best a problem with a solution, rather than a non-problem.

Let us return, then, to Judith Jarvis Thomson, who according to the constitutive hedonistic utilitarian has written and published several books and many articles whose subject matter is the property of failing to maximize the net long-run balance of pleasure over pain, but who has apparently been deeply confused about this property, about what instantiates it, and about reasonable ways of finding out what instantiates it. In the last section I focused on the very direct intuitive sense that constitutive hedonistic utilitarianism makes 'very poor sense' of Thomson's thoughts and practices – an intuitive sense which you may or may not share. But it is now time to sharpen that thought.

The problem is *not* simply that Thomson does not accept the analysis of 'wrong' that is given by constitutive hedonistic utilitarianism. Clearly, if you think, as Bentham did, that the definitions of 'right' and 'wrong' characterize not just what everyone is talking about, but what speakers *mean* to be talking about, then that would be pretty good evidence that this account does not apply to Thomson. And perhaps earlier

---

[3] More on these answers later.

4

generations of philosophers who were moved by something like the Common Subject Problem saw it in just these terms.[4] But an important part of the point that I aim to be making in this paper is that the Common Subject Problem survives this simple picture of what is entailed by a philosophical analysis. Indeed, since any philosophical analysis can be rejected, if the Common Subject Problem turns merely on the fact that Thomson rejects Bentham's analysis, Uniting would be hopeless.

Students of Rawls may identify the problem as securing a common *concept* for the subject matter of right and wrong – concepts being Rawls' tool for making sense of what guarantees different speakers and thinkers a common subject matter even when then disagree deeply in their *conceptions* about what fulfills this concept.[5] But even if Bentham and Thomson share some concept of right and wrong, that is not enough, I will be arguing, to make the problem go away. For the problem arises precisely because Thomson's views about what is right and what is wrong and the methods by which she forms those views are in such poor alignment with what actually makes things right and wrong, on the assumption of constitutive hedonistic utilitarianism.

Nor in any case, I believe, should we grant Rawls's idea that whenever speakers and thinkers share a common subject matter, it is necessary for them to share some common concept, in Rawls' sense, of what is at stake. It is true that the methodologically right way for any two conversational participants to make sure that their disagreement is substantive rather than merely verbal is to find some fixed point on which they can agree. This, I take it, is the kernel of truth in Rawls' talk of this distinction. But it does not follow from this that there must be any fixed point on which everyone can agree, for not everyone is conversing at once, and each of us might fruitfully use different fixed points when triangulating with different conversational partners. So even if there is no common concept of right and wrong on which everyone can agree, there is no deep problem here, for ethics.

It is also not sufficient to secure a common concept for all thinkers. If those thinkers are to avoid talking past one another, then the common concept employed by everyone thinking about 'wrong' must secure a common referent for such thought and talk as well. And so independently of the question of what their shared concept is, there is still a question of what this shared referent is. For example, Ralph Wedgwood [2007] has defended an account of the concept *ought* that can be shared by speakers with any views about what people ought to do and why, because no truth-conditions are built into the concept. But even for

<hr>

[4] Compare Ayer 1936.
[5] Rawls 1971, 5. Rawls credits the idea behind this distinction to Hart 1961.

Wedgwood, concepts constrain reference, and so there must still be a common referent secured by this shared concept, and our question is what that is.

Nor is the problem simply that Thomson has false beliefs – or even many false beliefs – about the referents of the words 'right' and 'wrong'. The thought that this is all that lies behind the intuition that constitutive hedonistic utilitarianism makes poor sense of Judith Jarvis Thomson is one of the thoughts motivating naïve Deflaters. Yet for most of human history, most people had false beliefs about the referent of the word 'water'. It was widely believed to be an element, for example, and to be infinitely divisible. So we know that false beliefs – even central false beliefs – are not by themselves any obstacle to something being the correct account of the subject matter of someone's thoughts and discourse.

False beliefs about the nature of water did not prevent our ancestors from correctly identifying water in their environment. Oh, they sometimes made mistakes, or even disagreed about what was water and what was not. When some of them encountered snow for the first time, they did not realize that it was water. Some of them drank clear spirits by accident, mistaking them for water. And some of them disagreed about whether clouds were made of water, or just made water. But these were all reasonable mistakes for them to make – the kinds of mistakes that someone will be led to make, who has a pretty reliable but imperfect way of forming beliefs about water in their environment, which might nevertheless misfire in unusual circumstances. For the most part, they were successful identifying water, because their way of identifying water was not affected by their false beliefs about its nature.

Judith Jarvis Thomson's beliefs about what is wrong are not like that. Not only do Thomson's moral beliefs turn out to be pervasively mistaken, her methods of determining what is wrong appear to be really horrible guides. According to constitutive hedonistic utilitarianism, what is wrong always depends on the consequences of the action, not only in the near-term, but in the long-term, as well. Yet Thomson routinely disregards the consequences of an action entirely when she considers whether it is wrong, only considering them in special cases (and Kant ignores them entirely!). So it seems to follow from constitutive hedonistic utilitarianism that Thomson's ways of forming moral beliefs are a really terrible guide to the truth.

This is a problem because Judith Jarvis Thomson undeniably *does* have a lot of knowledge – independent, apparently autonomous knowledge – about what is right and wrong. She knows, for example, that it is generally wrong to kill innocent people and to fail to keep one's promises. She knows that you are more likely to do the right thing if you are self-reflective about your behavior, that Hitler was evil, and that it is not wrong to proportionally harm a malevolent threat in self-defense. The more detail that we go into, the less that we will agree about exactly *what* Thomson knows, but we should all agree that she knows a fair bit.

And I venture to assert that she does not know these things by testimony. Whatever the possibility or virtues of relying on moral testimony more generally, Thomson, I venture, has not relied on testimony for everything that she knows about what is right and wrong. She has figured many of these things out on her own, or at best with some guidance from others about where to look for answers. And so her ways of discovering the answers to moral questions *must* be good enough to give her knowledge, at least in these cases. But it is just not obvious, if constitutive hedonistic utilitarianism is true, how this could be so.

This, I think, is the essence of the Common Subject Problem for ethics. The problem is what the common subject matter of ethics could be, such that the mistakes that various speakers and thinkers make about it could be intelligible mistakes *about it* – natural consequences of mistakes of ordinary human reasoning, or failures of a method that is reliable enough in its core cases to provide as much knowledge as a wide range of human moral thinkers actually do have, without attributing that knowledge to testimony.

It is this problem which, at bottom, lies at the heart of the motivations of contextualists and relativists in metaethics, and is one of the central attractions of expressivism as well. All of these views are natural manifestations of Dividing – contextualism its purest form, and relativism and expressivism some of its most prominent variants. All of these views deny that there is a common subject matter in my narrow sense for talk and thought about what is right or wrong – either because there is no common subject matter at all, or because the subject matter of ethics is so different in kind from other paradigmatic subject matters. And they are joined by a class of sophisticated new theories which mix the denial that there is a common subject matter with a second, expressive, implicated, or presupposed, aspect of moral thought and talk, which helps to explain the illusion that there was ever a common referent to 'right' and 'wrong' talk and thought.[6] Not all metaethical theories are motivated by this problem, and not all are obviously associated with a distinctive kind of answer to it. But many are.

The Common Subject Problem may have an answer – and indeed, it may even be, for all that I have said so far, that similar strategies are available for solving it to any answer as to what the common subject matter is, of ethical talk and thought. My goal has not been to refute the Deflater. But I hope to have said enough to show that *naïve* Deflating rests on an oversight. There *is* a puzzle, here, that deserves an answer. That much, at least, the Dividers are right about.

---

[6] For example, Finlay 2014, Perl forthcoming.

3   Strategies

So far, I have been focused on persuading Deflaters that there is at least something to puzzle about, in the neighborhood of the Common Subject Problem. If you are a Divider, you have been nodding your head along, or perhaps impatient for me to get to the point. But it is important for Dividers, too, to get clearer on what gives the Common Subject Problem its force. For once we give up the idea that a solution to the Common Subject Problem must secure an answer on which all competent thinkers could agree, it can no longer be obvious that it cannot have a solution.

And despite the fact, as I have argued, that the Common Subject Problem is a real problem, deserving of a real solution, there *are*, as I will now turn to show, a range of strategies that can be employed by a sophisticated Deflater, in order to answer the force of the challenge. And this means that naïve Dividing, like naïve Deflating, cannot be exactly right. Just as Deflaters need to take the problem seriously and provide it with real answers, Dividers need to take those answers seriously, on their own terms.

The first strategy available to the Deflater, I call *Narrowing*. The problem, as I have articulated it, stems from the fact that Judith Jarvis Thomson has a wide range of moral knowledge. The Narrower challenges the breadth of this range. Certainly, the Narrower admits, Thomson has some moral knowledge. But perhaps it is not as much as I might suppose. For example, maybe Thomson does not really know that it is *always* permissible to harm a malevolent threat in self-defense, but only that it is *generically* permissible to do so. And maybe she knows that Hitler was evil, but not that she was wronged by her cousin last week. The narrower the range of things that Thomson can be said to know, the less that it will take for her to be able to know them, and the weaker the things that Thomson can be said to know, the more ways of coming to know them will be sufficiently safe or reliable.

Narrowing the range of moral knowledge does not eliminate the problem all by itself. But it can be combined with other strategies. One particularly natural partner for Narrowing is what I call *Listening*. The Listener grants that the vast majority of Thomson's moral beliefs have not been formed on the basis of testimony – after all, she has both highly theoretical beliefs shared by no one else and highly specific judgments about examples that she was the very first to consider. But according to the Listener, these examples do not show that she has not inherited much of her moral knowledge by testimony, because most of our agreed-upon moral knowledge can be gained by Listening to the accumulated wisdom of our ancestors, who figured out that killing innocents is generically wrong, that it is generically right to keep one's promises, and the like.[7] According to the Listener, it is not so implausible that Thomson and many like her have

---

[7] Compare especially Sidgwick, in book IV of *The Methods of Ethics*.

knowledge that is either testimonial or at least broadly similar to ordinary testimony in its epistemological properties, with respect to the kinds of things that she was, indeed, explicitly taught by her parents and other elders when she was young, or with which she was enculturated in some other way by her community.

Listening is, as I have noted, naturally combined with Narrowing. The less that Thomson knows, the more of it can plausibly be attributed to the wisdom of her elders. In particular, the more of Thomson's knowledge can be claimed to be merely generic and shared with nearly everyone, the more of it can defensibly be claimed to have a broadly testimonial source. But while Listening is a promising strategy to account for highly general moral knowledge, it is not a very good strategy for accounting for knowledge of particular moral truths. Knowledge that the second Gulf War was wrong, for example, or that what Saanvi just did to you was wrong, are not the sorts of things that have been passed down from our ancestors. And they don't follow, either, from the wisdom of our ancestors, together with knowledge of particular facts of the situation.

At best, most of the wisdom of our ancestors is generic in flavor, rather than universal, and no particular claim ever follows from a generic. And even if some of the wisdom of our ancestors is completely universal, it will be too specific to encompass the complexities of the Gulf War or whatever Saanvi did to you. Yet it is too much to hope that Narrowing can plausibly eliminate all such particular moral knowledge. So the combination of Narrowing and Listening, I suggest, while they show that it is premature to conclude that Dividing is obvious in any way, is not enough to make the problem completely go away.

The Deflater has other available strategies, however. Some Deflaters will resort to the *Fine Individuation* of methods. Though the way W in which Thomson goes about acquiring moral beliefs, the Fine Individuator allows, is unreliable *in general*, there are several subject matters S such that applying W to S *is* a reliable way of acquiring answers to beliefs about S. So, for example, maybe Thomson's method is an unreliable way of forming moral beliefs about whether it would be wrong of Henry Fonda not to lay his cool hand across her brow when he happens to be across the room, or about whether it would be wrong to kill an innocent threat in self-defense. But it could still be a reliable way of forming moral beliefs about whether it would be wrong to unplug from the famous violinist, or wrong to kill a malicious threat. If this is so, then it could be that with a lot of help from the right answer to the generality problem for reliabilism, Thomson could end up having what it takes to have knowledge about the famous violinist and the malicious threat, even though she is in such a bad way with respect to Henry Fonda and the innocent threat (for example). And this is simply because *W-as-applied-to-S* is a highly reliable method, even though *W-as-applied-to-T* and *W-as-applied-to-R* are highly unreliable.

I don't believe, however, that Fine Individuation is a very promising strategy. In addition to the general problem that it relies on very specific commitments about the significance of reliability in

9

epistemology[8], it suffers from the problem that Thomson's way of forming moral beliefs is much too holistic, in order for us to be able to distinguish the method finely enough, in this way. Thomson's judgment that it is permissible to harm the malevolent threat, for example, is something that she does find intuitive prior to other judgments, but she comes to fully accept it only because it fits into a broader network of claims that she accepts which provide it a fuller explanation, including that it is permissible to kill even innocent threats, and that it is permissible for third parties to intervene as defenders.

Even more broadly, I've here been making references to some of the prominent claims made by Thomson in two of her justly famous papers about abortion and about self-defense, but even these two papers do not employ their methods separately; Thomson's explanation in 'A Defense of Abortion' [1971] for why it is permissible for a physician to provide an abortion to a woman whose life is threatened by her pregnancy depends quite closely on her claims in her work on self-defense [1991] about why it is permissible for a third party to intervene in cases of self-defense, even against innocent threats. When all of these claims hang together and the method of figuring them each out closely involves figuring them all out, it will simply not be possible to separately individuate the methods applied to each.

A much more promising alternative to Fine Individuation is what I call *Splitting*. Whereas the Fine Individuator assumes that Thomson employs a fundamentally unified method for forming moral beliefs and carves this unified method up differently for different contexts of use, the Splitter contends that moral inquiry is fundamentally disunified, relying on a wide variety of emotions, stereotypes, paradigms, or other psychological processes. If there is no single method that Thomson employs in forming her moral beliefs, the Splitter contends, it could be that some of these processes are reliable despite the fact that many are not.

For example, according to Haidt and Graham's [2007] Five Foundations theory, most of human moral thinking is built on the foundations of five (or so) foundational ways of thinking about moral issues. Their chief candidates are care, fairness, loyalty, respect, and purity. Haidt and Graham are, in my terms, Splitters. A Splitter may use the Five Foundations theory in order to explain why Thomson has some knowledge despite great unreliability, on the grounds that one of her methods – care – is reliable when used in isolation. But she is very unreliable on the whole, because she also applies some of the other methods, which, according to constitutive hedonistic utilitarianism, are not so reliable.[9]

---

[8] Compare Conee and Feldman 1998.
[9] For an example of Splitting in the wild, see Sinhababu unpublished. Sinhababu argues for hedonism by arguing that ethical judgments are formed in many ways, but introspective judgments of the goodness of pleasures are reliable, though other methods of belief formation are not.

There is far more to be said about Splitting than I can say here in this paper, which has broader goals. Indeed, the question of whether it works, and which views, if any, about the common subject of ethics it can be successfully used to defend, is I think one of the most interesting and central questions in normative and metaethical theory. What I want to emphasize, here, is that Splitting turns on a psychological hypothesis about the underlying ways we have of forming moral judgments or intuitions, and that this has the consequence that each account of the property picked out by 'wrong' will require a *different* such hypothesis in order to be sustained.

For example, whereas the defender of constitutive hedonistic utilitarianism needs the psychological hypothesis that one of the factors influencing Thomson's moral beliefs is her sympathy with the pleasures and pains of others, the defender of constitutive universalizable maximism needs the psychological hypothesis that one of the factors influencing Bentham's moral beliefs is his sense, very roughly, of the universalizability of his maxims. It follows that even if similar methods can be used to defend any hypothesis about the common subject matter of ethics, not all of these defenses are on a par. Some views will be better candidates for the common subject matter of ethics. And this, of course, is the central claim of *Uniters*.

4   Uniting

If the Common Subject Problem is, as I have argued, a real problem in need of a solution, and some of the moves for solving it are, as I have argued, specific to our underlying assumptions about the common subject matter of ethics, then that raises to salience the possibility that some accounts of the common subject matter of ethics are better than others.

This is the view taken by Uniters, who hold that *some* accounts of the nature of rightness and wrongness – perhaps (but not necessarily) constitutive hedonist utilitarianism among them – are worse than others, or even outright fatally flawed – by the lights of making sense of how we could all have been talking and thinking about the very same thing. On the flip side, Uniters therefore also believe that some accounts of the common subject matter of ethics are *better* than others. And so Uniters are committed to what I call the *Common Subject Project* – the project of finding an account of the nature of rightness and wrongness, *inter alia*, that can best make sense of how we are all talking and thinking about the same thing.

Uniting is not a new idea. On the contrary, much traditional work at the boundary between metaethics and normative ethical theory can be seen as motivated by the goal of Uniting. Take, for example, the traditional ideal observer theory. According to one version of this theory, what it is to be wrong is to be

what the most well-informed, rational, and impartial observer would disapprove of.[10] The ideal observer theory is compatible with utilitarianism. It is compatible with Kantianism. It is compatible with contractualism. And it is compatible with Thomson's theory. Which of these – or other – general normative ethical theories turns out to be correct will turn on what would be disapproved of by the most well-informed, rational, and impartial observer.

This means that the ideal observer theory offers everyone something that they can, at least in principle, *agree* on – a sort of shared conceptual core to wrongness. This is exactly what a Uniter should have wanted, if we assume that the nature of wrongness must be *analytic* or a *conceptual truth* – something that is accessible to all rational, reflective, and competent thinkers. And so this is part of what made the ideal observer theory attractive to an earlier generation of philosophers who would more naturally have accepted the background assumption that the nature of wrongness must be a conceptual truth, or that thinkers who are genuinely thinking about wrongness must share some conceptual grasp of it in common.

But the purposes of Uniting look different to philosophers with different sets of theoretical commitments, and I have been suggesting throughout this paper that the Common Subject Problem does not turn on such a problematic assumption about the competencies of ordinary thinkers. Still, the ideal observer theory offers much to Uniters even with more relaxed commitments. One of its chief virtues is that it allows us to trace the mistakes made by proponents of divergent moral theories (Thomson among them) to a single, natural, mistake about what would be approved of by the most well-informed, rational, and impartial observer. Because slightly different assumptions about what would be approved of by such an observer can lead to drastically different conclusions about what is right or wrong, the ideal observer theory offers a picture on which ordinary moral thinkers like Thomson are thinking about *exactly* the right sort of thing to successfully track what is right or wrong – namely, the approval of an ideal observer – and are even *pretty good* at assessing what relatively informed, relatively rational, relatively impartial observers would approve of, but just go slightly astray over which of these observers is *most* informed, rational, and impartial.

The ideal observer theory, I have been suggesting, is a traditional manifestation of the urge to Unite. And some of its prominent virtues as a Uniting theory remain, even if we relax some of the traditional assumptions from the philosophy of language that may have lain behind the actual commitments of some of the traditional proponents of the ideal observer theory. Independently of whether the ideal observer theory is something that all competent speakers implicitly grasp, the Uniter suggests that it is at least a *better candidate* for what everyone is *trying* to track – not because it makes the views of disparate thinkers about what is wrong

---

[10] Compare Firth 1951.

more accurate, but because it can rationally reconstruct large inaccuracies about what is wrong as traceable to relatively small inaccuracies about which observer is most ideal, or what she would disapprove of.

On this picture, the ideal observer theory constitutes a quite different kind of strategy for responding to the Common Subject Problem from those canvassed in the last section. Whereas those strategies began as attempts to offer a one-size-fits all response to the Common Subject Problem, the Ideal Observer Theory begins life as a very special hypothesis about the nature of that common subject matter. Its underlying strategy is not to Narrow knowledge, expand Listening, or Finely Individuate or Split the methods of moral belief formation so as to make some of them count as more accurate than others. Rather, it is to show how once we see the shape of the true common subject matter of ethics, the wide divergences that we see are just what we should expect, given a set of much *smaller* mistakes that do not impugn reliability.[11]

But the ideal observer theory is not alone, in being motivated by a similar conception of the project of Uniting. Another prominent contemporary form of Uniting is the view that has come to be known as Agent-Relative Consequentialism.[12] Agent-Relative Consequentialism is the thesis that what it is for an action to be wrong, is for it to be one that fails to maximize the agent-relative good, relative to its agent. Agent-Relative Consequentialism is perfectly consistent with traditional hedonistic act utilitarianism, because it could be that what is good-relative to each agent is always the same, and depends only on the long-term net balance of pleasure over pain in the universe. But it is also perfectly consistent with Thomson's rights-based deontological view, because it could be that violating a right is deeply agent-relatively bad relative to the agent who violates that right. Just as we can plug in different views about what the most informed, rational, and impartial observer would disapprove of into the ideal observer theory in order to get different results, similarly we can plug different views about what is agent-relatively good relative to whom, and to what degree, into Agent-Relative Consequentialism, in order to get different results.

Agent-Relative Consequentialists often make very general claims about just how many different moral theories can be captured as equivalent to Agent-Relative Consequentialism plus special commitments about what is good relative to whom.[13] And there are multiple purposes to which such claims might be put. For example, if we can first prove that all moral theories are equivalent to Agent-Relative Consequentialist plus certain assumptions about what is good relative to whom, and then we can prove something general about Agent-Relative Consequentialism, then we can in turn draw a general conclusion about all moral

---

[11] Accepting the ideal observer theory is, of course, compatible with *also* appealing to the strategies of Narrowing, Listening, Fine Individuation, and Splitting as additional resources to make sense of how Thomson and others know what they do.
[12] Among other sources, Dreier 1993, Louise 2004, Portmore 2005.
[13] Compare especially Brown 2011.

theories, independently of the truth of Agent-Relative Consequentialism. Some discussions of Agent-Relative Consequentialism ve this flavor.[14] But most proponents of Agent-Relative Consequentialism want to say more – they want to say that Agent-Relative Consequentialism in some way reveals the *nature* of what is right and wrong, and that part of what makes it a good account is that it is in some way neutral between more substantive commitments. This motivation is a straightforward manifestation of the impulse to Unite.

Another common contemporary perspective in both metaethics and normative ethical theory is that moral rightness and wrongness are determined by the *balance of reasons*. This view, which traces its lineage to Ross's (1930) theory of *prima facie duties*, can be motivated independently of the impulse to Unite. But Uniters may be particularly attracted to the Balance of Reasons theory. If the Balance of Reasons theory is right, then even though we may disagree sharply in many cases about what is right and wrong, and hence most of us are mistaken much of the time about particular cases, these mistakes are the natural result of smaller mistakes about the balance of reasons. We may be pretty good at identifying what is a reason to do what, and even at identifying roughly how strong a reason it is, but just less good at comparing the strengths of reasons in close or difficult cases, or when many reasons combine. This would explain why we get things right in easy cases – cases where there are no deep conflicts – but struggle in cases of competing reasons.

As we've seen throughout this paper, Uniters who endorse theories like the ideal observer theory, agent-relative consequentialism, or the balance of reasons theory must motivate their views on two fronts. They will push back against Deflaters in ways that we have already explored in our statement of the Common Subject Problem in section 2. And they will push back against Dividers by defending the claim that their candidates are genuinely better candidates for the referent of 'wrong'.

Dividers, of course, will resist this, regardless of which sort of view – agent-relative consequentialism, balance of reasons theory, or ideal observer theory – Uniters appeal to. In earlier generations, philosophers thought that to solve the Common Subject Problem we must be able to articulate a concept definition or at least a core constraint on a concept that every competent thinker can accept. On such a conception of the Common Subject Problem, I think that we should agree that the problem will generalize, in either or both of two ways – which we can illustrate with the example of ideal observer theory. First, since philosophers can disagree about anything, the ideal observer theory is no closer to avoiding disagreement than constitutive utilitarianism. And second, the ideal observer theory only reduces mistakes about what is wrong to smaller mistakes about which observer is most ideal, if we have a common subject matter for talk about what is 'ideal'. And this seems somehow only to push the problem back, rather than solving it.

---

[14] Compare Ross 2006.

But I have been endeavoring to resist this way of conceiving of the Common Subject Problem. The force of the Common Subject Problem, as I have articulated it, is to explain how people can each attain *knowledge* about the property of wrongness in spite of their differences in what they believe about it and how those beliefs affect how they inquire about it. So the answer is that given this new conception of the problem, Uniting views do not need to tell us something on which thinkers can all agree. They just need to assume that there is actually something in common to how people go about gaining moral knowledge – and their account of the common subject matter of talk about what is wrong must be tied to this.

Take, for example, the case of the ideal observer theory. This theory can explain how different people can come to each have knowledge about what is wrong despite very different views about the extent and nature of wrongness, so long as when people try to figure out what is wrong, one of the things that they try to do is to think objectively and rationally. When they do so, according to the ideal observer theory, they are not self-consciously *applying* the ideal observer theory, but they *are* occupying the standpoint of an ideal observer as well as they are able. Indeed, this is precisely how Hume's own ideal observer theory works – we don't learn about virtue by thinking about what an impartial observer would approve of, but by approving of things when we are as impartial as we can get ourselves to be.

Since the ideal observer theory does not require the assumption that thinkers all realize that the ideal observer theory is true, but only that they think in ways that are mapped by the ideal observer theory, it is no problem that not everyone agrees that the ideal observer theory is true. What *would* be a problem, is if there are people who do not try to be better-informed, more rational, or more impartial when they think about what is wrong. But I conjecture that it is more plausible that the strategies of Narrowing, Listening, Fine Individuation, and Splitting can explain the extent of such a person's moral knowledge than that they could suffice to explain Thomson's, if constitutive hedonistic utilitarianism were true. And since thinkers are not required to think *in terms* of ideality of observers by this view, no regress is threatened by acknowledging that there may also be a separate common subject problem about ideality itself – also amenable to Uniting solutions of its own.

In this section I've been trying to show that Uniting is not in any way a novel idea in response to a strange new problem, but rather is a theoretical impulse deeply embedded into both metaethics and normative ethical theorizing at the most general level. If we fail to understand the way in which these different highly general moral theories are motivated by the project of Uniting, then we miss something central about their appeal, render invisible something important that such views have in common with one another, and fail to grasp one of the most interesting places where metaethical inquiry and normative ethical theory hit the pavement in tandem.

5 Distinguishing

As I have been keen to emphasize, the Common Subject Problem is the problem of how anything could be the common subject matter of moral thought and talk. Solving it requires defending some candidate view about the nature of that common subject matter. As we have seen, it is not a problem exclusively for reductive metaethical views, but it is true that it is a problem *especially* for those views. One would expect that if this problem is really central to contemporary metaethics and normative ethical theory, then one would see at least some discussion of this problem, at least as an objection to reductive realism in metaethics. But in fact, we see very little such discussion in contemporary literature.

Indeed, with very few exceptions, almost all discussion of reductive realism in metaethics takes completely for granted that it doesn't matter *at all* which descriptive or natural property moral wrongness turns out to be. What is much more often discussed, is the idea that synthetic reductive realists in metaethics – a subclass of those who face the Common Subject Problem – face a special problem accounting for *how* moral words could have come to pick out this property, whatever it is. This putative problem is known as the Moral Twin Earth problem, after the class of examples used by its proponents, Terry Horgan and Mark Timmons, in order to introduce it.[15]

Horgan and Timmons' discussions of Moral Twin Earth are important and have focused attention on important and previously neglected issues. But as I will now argue, insofar as the Moral Twin Earth Problem has force as a general problem for reductive views in metaethics, that is because it gets at same kinds of issues raised by the Common Subject Problem. It is a consequence of this that discussion of the Moral Twin Earth problem has *distorted* our understanding of the issues lying behind the Common Subject Problem. By misrepresenting the nature of the problem, it has led to misunderstanding both the scope of the problem and the space of possible answers.

According to Horgan and Timmons, though there is no special independent problem about what property we could all be talking about, there *is* a special problem about how we could all be talking about it.

---

[15] Horgan and Timmons first laid out the Moral Twin Earth problem in their 1991 and 1992, but have followed up with additional papers extending the reasoning to alternative accounts of content determination (1996a and 2009), responding to critics (2000 and 2015), and most helpfully, by doing more to lay out the structure of the reasoning that they take to be active in the argument (1996b). Responses have been published by, among others, Geirsson 2003, 2005, 2014, Rubin 2008, 2014a, 2014b, Copp 2000, Merli 2002, Gert 2006, van Roojen 2006, Sonderholm 2013, Freiman 2014, Dowell 2016, Dunaway and McPherson 2016, and Williams forthcoming. Most of these responses explore a relatively small region of the space of possible theories of content-determination.

The problem, *according to them*, is not in first-order moral semantics, but in *metasemantics*, or the theory of content determination:[16]

	Moral Twin Earth Problem	Common Subject Problem
Is there a problem with identifying a property that we are all talking about?	No	Yes
Is there a problem with explaining how we could be talking about it?	Yes	No

The way that Horgan and Timmons see things, it doesn't matter whether there is a plausible candidate for what property everyone could be talking and thinking about, when they are talking and thinking about what is wrong. Indeed, their argument works by *granting at the outset* that there is such a property picked out by 'wrong' on Earth. The problem, they claim, is that no possible theory of content determination could explain how we are all talking about that property, because any such theory will have counterexamples at other possible worlds.

Similar claims are made by others. For example, Bart Streumer [2017] also argues that the fundamental problem that rules out reductive realism about the normative is that there can be no adequate answer to what makes it the case that normative predicates ascribe the properties that they do. Indeed, it has come to be part of the common lore in metaethics that reductive realists have a special problem in metasemantics. But as I will now explain, this is simply a mistake. And it is a mistake that has distorted our understanding of both the scope and range of solutions to what is in fact a very real problem.

The way that Horgan and Timmons argue is simple. They first stipulate, for *reductio*, that on earth we pick out a "consequentialist property" with the word 'wrong'. And then they consider a possible theory about in virtue of what 'wrong' picks out this consequentialist property – in their original treatment, they started with Richard Boyd's [1988] theory, which has consequently gotten the most discussion, but they have also considered other example theories. Then they construct a proposed counterexample – called 'moral twin earth', in which Boyd's theory arguably predicts that people use the word 'wrong' to pick out a "deontological property", but intuitively speakers mean the same thing as us by 'wrong'. Next they argue that they will be able to construct similar counterexamples to *any* other naturalistic theory of content determination. And

---

[16] We might put this point slightly more carefully by saying that although Horgan and Timmons do believe that there is no candidate for reductive naturalists can take 'wrong' to refer to, they believe that the problem is not with the candidates *per se*, but with the fact that no theory of reference determination could have made these what we are actually talking about.

finally, they assume without argument that their choice of which property they stipulated for *reductio* is picked out by 'wrong' is immaterial to the argument.

Not surprisingly, my diagnosis is that on the contrary this stipulation *is* material to the argument, and indeed that it is essential. The evidence is both circumstantial and direct. Taking the circumstantial evidence first, the Common Subject Problem is clearest and most familiar when we try to imagine how Bentham and Thomson could both be talking either about the property of failing to maximize the long-term balance of pleasure over pain. Horgan and Timmons' stipulations that Earthlings are talking about a "consequentialist property" and twin-earthlings about a "deontological property" sets us up to think about the relationship between earthlings and twin-earthlings as a special case of the way that we think about the relationship between Bentham and Thomson. The point of the example is to prime the central worry of the Common Subject Problem.

But we can also get much more direct evidence that this must be what is going on, by seeing just how ill-suited the structure of the argument is, at establishing what Horgan and Timmons say explicitly that it is supposed to establish. Many philosophers have published responses to Horgan and Timmons, and most of these responses are preoccupied with assessing whether Horgan and Timmons' moral twin earth example is a genuine counterexample to Boyd's theory of content determination, or whether it or similar counterexamples will work against other, similar, theories of content determination, or to the task of constructing a theory of content determination that is immune to similar counterexamples.[17] This work is all very interesting and there is much to be learned from it, but the fact that there is so much of it tends to obscure the most fundamentally important fact about Horgan and Timmons' argument, which is that it does not even constitute a *prima facie* challenge unless there can be some ground to at least suspect that we will be able to construct intuitive counterexamples to *any possible theory* of content determination – the whole force of the argument turns on the strength of their evidence for this universal quantifier.

With some perspective, however, we can see that this is a strange allegation indeed. Horgan and Timmons claim to know how to construct counterexamples to all possible theories, and so by implication they claim to know which features of cases our intuitions about content (mediated by intuitions about agreement or disagreement) are sensitive to. On an intuitive gloss, what they are thinking is that our intuitions about content will track the motivational role that words play. But if we know which features of cases our intuitions about content are sensitive to, then we know which features to incorporate into a theory of content determination, in order to avoid counterexamples.

---

[17] Compare note 13.

So, in particular, if we know that intuitions about content are sensitive to motivational role, then all that we have to do is to incorporate motivational roles into our theory of content determination. But since motivational rules are natural, there is no in-principle obstacle to incorporating them into a naturalistic theory of content determination. So it simply doesn't make sense to be confident that you can construct a counterexample to *any* theory of how a term gets to refer to a particular property unless you are antecedently confident that the term does not pick out that property at all. Otherwise, you are simply playing by different rules than you are allowing your opponent.

Compare a simple example. Suppose that Horgan and Timmons sought to show that there can be no naturalistic theory of donuts, by appealing to counterexamples involving holes, but insisted that naturalists are not allowed to mention holes in their theory. That would be cheating. If the counterexamples involve holes, then a theory can be constructed to avoid the counterexamples by appealing to holes in the theory. Since holes are natural, a naturalistic theory of donuts is allowed to appeal to them. Similarly, since motivational roles are natural, a naturalistic theory of content determination is allowed to appeal to them.

I conclude not only that Horgan and Timmons have not *shown* that their argument works, but that we can directly read off from the structure of their argument that no such argument can possibly work. Any reason to be confident that we know enough about what intuitions about content are sensitive to in order to go on constructing counterexamples to new theories will be a reason to be confident that we can construct a theory to get around them.[18]

Of course, it is consistent with this observation that a theory of content determination must be sensitive to *different things* in order to adequately capture the reference of normative words like 'wrong' than it needs to be sensitive to in order to adequately capture the reference of non-normative words like 'blue' and 'Madagascar'. So I do not intend to be suggesting that the moral twin earth examples are not interesting for the theory of content determination. Indeed, reductive realists in metaethics may even carry some burden that is not carried by, for example, contextualists or expressivists, because moral twin earth examples commit them to more unorthodox views in the theory of content determination.[19] So Moral Twin Earth cases could

---

[18] Horgan and Timmons do claim to have a recipe for constructing counterexamples to any possible theory – one that builds in a lot of controversial theory. They assume that if the question of which property is picked out by 'wrong' is a synthetic one, rather than an analytic one, then it must be *a posteriori*, that if it is *a posteriori*, then learning it must involve ruling out some possibility, and that the possibility that is ruled out must be a genuine metaphysical possibility. Finally, they assume that when we consider this possibility, we will find it to be an intuitive counterexample to this theory of content determination. Though it would take us too far astray, I believe that this reasoning goes wrong at every step. We should expect Horgan and Timmons' counterexamples to stand on their own.

[19] For the record, my own view is that this is not likely to be a serious cost at all. Most familiar theories of content determination, including causal, co-variation, informational, teleological, and inferential theories, are focused on the contents of *beliefs*. But beliefs are just one propositional attitude among others, and there is no obvious *a priori* reason why their contents should be privileged

be quite interesting and important for the theory of content determination, quite independently of its interest for metaethics.

And if what I have argued is correct, then on any charitable reading, Horgan and Timmons were latching onto something important in metaethics, as well – the Common Subject Problem. But their way of getting us to think about it was imperfectly conceptualized. We would do better to drop it and focus on what it takes to identify a common property, setting metasemantics aside. If what I have argued is correct, then there are – indeed, there must be – adequate answers to what metasemantic theory can avoid the Moral Twin Earth problem. But proponents of such a theory still face the Common Subject Problem.

7   Summing Up

My aim throughout has been to conceptualize not a new problem, but rather, an old and familiar problem, within both metaethics and normative ethical theory, which has nevertheless been but dimly understood, even by those who seem to take it most seriously. I have not tried to argue that this problem can be solved or that it cannot. What I have endeavored to do, is to convince Deflaters that there is a genuine issue here, and to convince Dividers that there are a range of strategies for response that merit fruitful engagement. And I have tried to provide circumstantial evidence to everyone that this problem has played a central role at the intersection of metaethics and normative ethical theory, and like other such problems, should be treated as something that we can recognize others disagree about, even when we ourselves find the answers obvious.

I have also tried to offer one strand of a diagnosis of why this problem has seemed so particularly hard to grasp in the last thirty years, for the literature that has brought it into force most clearly has misconstrued a problem about semantics as one about metasemantics. This misconstrual has had the unfortunate effect of writing the strategy of Uniting out of visibility. But if I am right, then perhaps Uniting is an intelligible strategy and interesting project after all. Perhaps it is no more the case that we and moral twin earthers are all thinking about a consequentialist property than it is that we are all thinking about a consequentialist property, but still quite possible that we all – moral twin earthers included – are thinking about what is right and wrong.[20]

---

over the contents of desires, hopes, wishes, or intentions. So offering a larger place for the role of desires and other attitudes in a theory of content determination is independently motivated out of considerations of non-arbitrariness. Yet desires and other attitudes are precisely what are shifted in twin earth style cases.

MARK SCHROEDER

University of Southern California

maschroe@usc.edu

References

Boyd, Richard [1988]. 'How to Be a Moral Realist.' In Geoff Sayre-McCord, ed., *Essays on Moral Realism.* Ithaca: Cornell University Press, 181-228.

Brown, Campbell [2011]. 'Consequentialize This.' *Ethics* 121(4): 749-771.

Conee, Earl, and Richard Feldman [1998]. 'The Generality Problem for Reliabilism.' *Philosophical Studies* 89(1): 1-29.

Copp, David [2000]. 'Milk, Honey, and the Good Life on Twin Earth.' *Synthese* 124(1-2): 113-137.

Dowell, Janice [2016]. 'The Metaethical Insignificance of Moral Twin Earth.' *Oxford Studies in Metaethics* 11: 1-27.

Dreier, James [1993]. 'The Structure of Normative Theories.' *The Monist* 76(1): 22-40.

Dunaway, Billy, and Tristram McPherson [2016]. 'Reference Magnetism as a Solution to the Moral Twin Earth Problem.' *Ergo* 3(25): 639-679.

Finlay, Stephen [2014]. *A Confusion of Tongues.* Oxford: Oxford University Press.

Firth, Roderick [1951]. 'Ethical Absolutism and the Ideal Observer.' *Philosophy and Phenomenological Research* 12(3): 317-345.

Freiman, Christopher [2014]. 'Goodness and Twin Earth.' *Erkenntnis* 79(2): 445-460.

Geirsson, Heimir [2003]. 'Moral Twin Earth: The Intuitive Argument.' *Southwest Philosophy Review* 19(1): 115-124.

_____ [2005]. 'Moral Twin Earth and Semantic Moral Realism.' *Synthese* 62(3): 353-378.

_____ [2014]. 'Moral Twin Earth, Intuitions, and Kind Terms.' *Croatian Journal of Philosophy* 14(1): 91-110.

Gert, Joshua [2006]. 'Problems for Moral Twin Earth Arguments.' *Synthese* 150(2): 171-183.

Haidt, Jonathan, and Jesse Graham [2007]. *Social Justice Research* 20(1): 98-116.

Hart, H.L.A. [1961]. *The Concept of Law.* Oxford: Clarendon Press.

Horgan, Terence, and Mark Timmons [1991]. 'New Wave Moral Realism Meets Moral Twin Earth.' *Journal of Philosophical Research* 16: 447-465.

_____ [1992]. 'Troubles for New Wave Moral Semantics: The Open Question Argument Revived.' *Philosophical Papers* 21(3): 153-175.

_____ [1996a]. 'Troubles for Michael Smith's Metaethical Rationalism.' *Philosophical Papers* 25(3): 203-231.

_____ [1996b]. 'From Moral Realism to Moral Relativism in One Easy Step.' *Critica: Revista Hisponoamericano de Filosofía* 28(83): 3-39.

_____ [2000]. 'Copping Out on Moral Twin Earth.' *Synthese* 124(1-2): 139-152.

_____ [2009]. 'Analytical Functionalism Meets Moral Twin Earth.' In Ian Ravenscroft, ed., *Mind, Ethics, and Conditionals: Themes from the Philosophy of Frank Jackson.* Oxford: Oxford University Press.

_____ [2015]. 'Exploring Intuitions on Moral Twin Earth: A Reply to Sonderholm.' *Theoria* 81(4): 355-375.

Louise, Jennie [2004]. 'Relativity of Value and the Consequentialist Umbrella.' *The Philosophical Quarterly* 54(217): 518-536.

Merli, David [2002]. 'Return to Moral Twin Earth.' *Canadian Journal of Philosophy* 32(2): 207-240.

Moore, G.E. [1903]. *Principia Ethica.* Cambridge: Cambridge University Press.

Perl, Caleb [forthcoming]. 'A User's Guide to Hybrid Tools.' Forthcoming in *Mind.*

Portmore, Douglas [2005]. 'Combining Teleological Ethics with Evaluator Relativism: A Promising Result.' *Pacific Philosophical Quarterly* 86(1): 95-113.

Rawls, John [1971]. *A Theory of Justice.* Cambridge: Harvard University Press.

Ross, Jacob [2006]. *Acceptance and Practical Reason.* PhD dissertation, Rutgers University.

Ross, W.D. [1930]. *The Right and the Good.* Oxford: Oxford University Press.

Rubin, Michael [2008]. 'Sound Intuitions on Moral Twin Earth.' *Philosophical Studies* 139(3): 307-327.

_____ [2014a]. 'Biting the Bullet on Moral Twin Earth.' *Philosophical Papers* 43(2): 285-309.

_____ [2014b]. 'On Two Responses to Moral Twin Earth.' *Theoria* 80(1): 26-43.

Sidgwick, Henry [1903]. *The Methods of Ethics*, 7ᵗʰ Edition. Indianapolis: Hackett.

Sinhababu, Neil [unpublished]. 'The Epistemic Argument for Hedonism.' Unpublished paper, available at https://philpapers.org/archive/SINTEA-3.pdf.

Sonderholm, Jorn [2013]. 'Unreliable Intuitions: A New Reply to the Moral Twin Earth Argument.' *Theoria* 79(1): 76-88.

Streumer, Bart [2017]. *Unbelievable Errors*. Oxford: Oxford University Press.

Thomson, Judith Jarvis [1971]. 'A Defense of Abortion.' *Philosophy and Public Affairs* 1(1): 47-66.

_____ [1991]. 'Self-Defense.' *Philosophy and Public Affairs* 20(4): 283-310.

van Roojen, Mark [2006]. 'Knowing Enough to Disagree: A New Response to the Moral Twin Earth Argument.' *Oxford Studies in Normative Ethics* 1: 161-194.

Wedgwood, Ralph [2007]. *The Nature of Normativity*. Oxford: Oxford University Press.

Williams, J.R.G. [forthcoming]. 'Normative Reference Magnets.' Forthcoming in *Philosophical Review*.