Why Moral Paradoxes Support Error Theory[1]
The Journal of Philosophy

Abstract: The aim of this article is to present a somewhat pessimistic picture of what moral reasoning can hope to achieve, and to draw a striking metaethical conclusion from it. The pessimistic picture emerges from the various puzzles, paradoxes and impossibility results that contemporary moral philosophers struggle with; for example, the axiological paradoxes first developed by Parfit. These puzzles and paradoxes show that some troublingly counter-intuitive moral claims *must* be true. Indeed rather a lot of them must be true. The striking metaethical conclusion is that this allows for a new and effective defence of moral error theory. Moral error theory is often thought to be problematic because of its counter-intuitive and troubling consequences for our ordinary 'first order' moral judgment. But the ubiquity of puzzles, paradoxes and impossibility results in ordinary moral philosophy shows that this is *everybody's* problem. And indeed, there is a sense in which the problem is worse for the opponents of error theory than for error theory.

Moral error theory is the view that no moral claim can be true and all are therefore mistaken.[2] While there is a compelling case to be made for this view, the conclusion is difficult to accept. Consider for example the claim that the Holocaust was morally problematic. If error theory is correct this claim is not true. It is hard to see how this could be and easy to think that anyone who takes the possibility seriously has lost their moorings in reality, is trying to be difficult, or something worse still. This simple but important worry is at the heart of much ordinary scepticism about error theory.[3] For this reason I refer to the challenge it presents as *the fundamental challenge* for error theorists.

[2] See for example Richard Joyce, *The Myth of Morality* (Cambridge: Cambridge University Press 2001); Jonas Olson, *Moral Error Theory: History, Critique, Defence* (Oxford: Oxford University Press, 2014); Bart Streumer, *Unbelievable Errors* (Oxford: Oxford University Press 2017); and Wouter Kalf,. *Moral Error Theory* (Palgrave MacMillan 2018). I shall focus exclusively on *moral* error theory and shall use 'error theory' to denote moral error theory. I shall not discuss metanormative error theories or whether the former imply the latter, for more on which see Richard Rowland, 2013, "Moral Error Theory and the Argument from Epistemic Reasons", *Journal of Ethics and Social Philosophy* 7(1): 1-24 (2013); Guy Fletcher, "Pain for the Moral Error Theory? A New Companions in Guilt Argument", *Australasian Journal of Philosophy* 96(3): 474-482 (2018); and Christopher Cowie, *Morality and Epistemic Judgment* (Oxford: Oxford University Press, 2020).

[3] See e.g. Ronald Dworkin, *Justice for Hedgehogs* (Boston, Mass.: Harvard University Press, 2011), Chs. 2-5; and David Enoch, *Taking Morality Seriously* (New York: Oxford University Press, 2011): 121. Enoch's presentation is naturally read as claiming that error theory's morally troubling consequences are *sufficient* for its rejection. An opponent of error theory could also make the weaker claim that they *count against* error theory though are not sufficient for its rejection. For critical discussion of the use of the sufficiency claim in the context of so-called 'Moorean' arguments, see e.g. Tristram McPherson, Moorean Arguments and Moral Revisionism. *Journal of Ethics and Social Philosophy* (2009), 2: 1-25..

Error theorists respond by denying that we should judge their theory in this way; we should set its implications for ordinary moral judgment to one side and instead judge it on its *metaethical* merits. This is often accompanied by some means of sweetening the pill. One such means is to appeal to *debunking* arguments; accounts of the origins of moral judgments – in terms of, for example, socio-cultural function – that explain why error theory's incompatibility with them *seems* (but isn't) troubling.[4] The effect of this is limited. I have yet to meet anyone who initially thought error theory seriously undermined by its incompatibility with (e.g.) the immorality of the Holocaust, yet subsequently revised this judgment on being provided with an account of the causal origins of that very judgment. Add to this that many contemporary non-error theorists willingly accept socio-cultural just-so-stories of moral judgment but see them as part of a *non*-debunking, even a *vindicatory*, supplement to their view.[5]

An alternative means of sweetening the pill is offered by fictionalists, conservationists and other 'insulationist' responses to the 'what now?' problem; the problem of how one ought to proceed if one has accepted error theory.[6] Insulationists claim that error theory does not entail that one ought to abandon ordinary moral judgments outside of the philosophy classroom. One can continue to judge that, for example, the Holocaust was morally problematic, in everyday contexts. This blunts the force of the fundamental challenge. But there are some problems here too. Firstly, *whatever* one's response to the 'what now?' problem, error theory still entails that it is not true that the Holocaust was morally problematic. There is no getting away from this. Secondly, it is simply not clear that one *can* insulate ordinary moral judgment from metaethics as fictionalists and conservationists recommend without either compromising the motivational efficacy of that judgment, or requiring that ordinary judgers either display irrationality or an unworkable form of cognitive dissonance.[7]

---

[4] The locus classicus is Joyce 2001, Ch. 6. See also Richard Joyce, *The Evolution of Morality*. (Boston, Mass.: MIT Press, 2006) and Olson 2014, Ch. 7. One could also read Streumer's (2013, 2017 Ch. 9) claim that it is impossible to believe error theory (Bart Streumer, "Can We Believe the Error Theory?" *Journal of Philosophy* 110(4): 194-212, (2013)) as a debunking argument, though of a rather different sort. An exegetical issue here is whether these error theorists see debunking arguments as a defence of error theory against the fundamental challenge, or as (part of) an argument for error theory. The simplest answer is 'both'. For example, Olson (2014) uses a debunking argument as a response to the 'Moorean' challenge that error theory is incompatible with the wrongness of torturing children for fun (140, 146). This is a defence against a specific form of the fundamental challenge. But he also describes it as 'the second step of the argument from queerness' (141) and uses it as part of a parsimony-based case against non-naturalism (147). This is part of an argument for error theory.

[5] For example, Sharon Street uses these arguments as the basis of an epistemological case for a broadly constructivist view. See Sharon Street, "A Darwinian Dilemma for Realist Theories of Value", *Philosophical Studies* 127: 109-166 (2006). 'Debunking' arguments are used as the basis of an epistemological case for realist views by e.g. David, Copp ("Darwinian Skepticism about Moral Realism" *Philosophical Issues* 18(1) 186-206 (2008)), David Enoch (2011) and Derek Parfit (*On What Matters vol. 2* (Oxford: Oxford University Press, 2011): Chs. 32-33). For further discussion see William FitzPatrick, "Debunking Evolutionary Debunking of Ethical Realism", *Philosophical Studies* 172: 883–904 (2015).

[6] The expression 'insulationism' is used in (e.g.) Diego Machuca, "Moral Skepticism, Fictionalism and Insulation", in D. Machuca (ed.) *Moral Skepticism: New Essays* (New York: Routledge, 2018). Forms of it include fictionalism (e.g. Joyce 2001) and conservationism (e.g. Olson 2014).

[7] For useful discussion see Hallvard Lillehammer, "Moral Error Theory", *Proceedings of the Aristotelian Society* 104(2): 93-109 (2004); and Matt Lutz, "The 'Now What?' Problem for Error Theory", *Philosophical Studies* 171(2): 351-371 (2014).

I provide a *new* response to the fundamental challenge for error theory. I argue that even if we allow that error theory's consequences for moral judgment *are* genuinely troubling and *can* legitimately be counted against it, the *alternative* metaethical theories in fact do not fare much better. They too have deeply counter-intuitive and troubling consequences for moral judgment. This should sound surprising. Consider, for example, realism; the view that there are stance-independent moral properties about which we can, and sometimes do, make correct claims.[8] It is not obvious that this theory has *any* moral consequences, let alone any as counter-intuitive and troubling as error theory. I argue otherwise. *Every* metaethical theory has troubling and counter-intuitive moral consequences in the relevant sense. So error theory should not be singled out for special punishment. My argument is based not in abstract considerations of high metaethics but rather in the comparatively homely field of ordinary moral philosophy.[9] I offer a pessimistic reading of ongoing work on paradoxes, puzzles and impossibility results in recent decades. This work shows that if moral claims can indeed be true – as all but error theorists accept - then we are unavoidably committed to accepting a non-negligible number of claims that are extremely counter-intuitive and troubling. I focus on two examples to make my point: trolley problems and axiology.

Let me state the point modestly. I do not claim to show that the consequences of error theory are *exactly as troubling as* those of the alternatives, like realism. Realistically, I think this is probably too ambitious. The consequences of error theory may be *a bit worse* than those of the alternatives. But 'a bit worse' is good enough for my purposes. Error theory undoubtedly fares well against its competitors - realism, expressivism, and so on - in many respects, for example metaphysics and epistemology. The respect in which it most obviously fares *badly* is with regard to its counter-intuitive consequences for moral judgment. My aim is to show that it doesn't fare *that much* worse in this respect; and certainly not badly enough to reject it out of hand. If I can show this, then error theory's superiority in other regards - metaphysics, epistemology and so on – will be enough to make it come out in 'first place' when the explanatory points get totted up at the end of the day.

I. The Consequences of Error Theory

If error theory is true, then no moral claim can be true and all are mistaken.[10] This contributes to scepticism about error theory. I do not attempt to debunk or downplay this. My aim is simply to show that the alternatives fare surprisingly little better. Let's start by getting clear on exactly what error theory's consequences for ordinary moral judgment actually are - because they are not quite as bad as one might at first think. Error theorists, as we have seen, are committed to the following:

(A) It is not true that the Holocaust was morally problematic.

Ordinarily, if someone were to say this, we would take them to be committed to:

---

[8] See e.g. Russ Shafer-Landau, *Moral Realism: A Defense* (New York: Oxford University Press 2003); Enoch 2011; Parfit 2011.

[9] I draw this distinction in order to distance my argument from the claim that realism is 'immoral' because it entails that the moral facts are a hostage to modal fortune: e.g. Mark Heller, "The Immorality of Modal Realism, Or: How I learned to Stop Worrying and Let the Children Drown", *Philosophical Studies* 114(1-2): 1-22 (2003). My argument is unrelated.

[10] I withhold judgment on whether the fact that error theory entails no moral claims can be true entails that all are false. For discussion see e.g. Wouter Kalf, "Moral Error Theory, Entailment and Presupposition", *Ethical Theory and Moral Practice* 16(5): 923-937 (2013); Kalf 2018; Caleb Perl and Mark Schroeder, "Attributing Error Without Taking a Stand", *Philosophical Studies* 176(6): 1453-1471 (2019).

(B) The Holocaust was morally unproblematic.

The latter, we might think, is an implication of the former. Importantly though, if error theory is correct then this natural thought is not true. If error theory is true, there are no moral properties at all. So (B) – which ascribes the property of being morally unproblematic - *couldn't* be true. So if error theory is true, (A) doesn't imply (B). The point isn't particular to judgments – like (A) and (B) - that ascribe the property of being morally (un)problematic. Consider as another example the deontic properties of being permissible, obligatory or forbidden. Error theory is incompatible with the truth of the proposition that the Holocaust was morally impermissible. But it does not thereby imply that it is true that the Holocaust was morally permissible. Indeed it implies that *this* is not true either.

This manoeuvre is important because it makes the consequences of error theory for ordinary moral judgment a lot less counterintuitive and troubling than they might at first seem. Error theorists don't regard the Holocaust as having some moral properties, namely 'good' or 'ok' ones. They regard nothing as having any moral properties at all.

This observation makes use of a feature of error theory that might cause one to worry about its coherence on purely formal grounds. If error theorists think that it is not true that the Holocaust was morally impermissible, mustn't they - as a matter of logic following form the principle known as 'duality' – accept the horrifying conclusion that it was morally permissible? This is the subject of much recent discussion.[11] There are a number of options for error theorists. One is to regard the principle of duality as a highly abstract first-order moral claim that, as such, is untrue.[12] Another is to understand error theory not as claiming that moral propositions are false, but as claiming that they suffer from a presupposition failure and so are neither true nor false. This allows error theorists to hold on to the view that no moral propositions are true, whilst not entailing that duality is false.[13] I shall assume in what follows that there is some such solution available to this formal challenge, though I shall not take a stand on which.

Why assume that there is some solution to be found? I am broadly in agreement with Streumer and Wodak's argument that formal objections to error theory would overgeneralise.[14] Consider spiritual purity. An error theory of judgments about this property is surely reasonable. But by analogy with the above objection to moral error theory, if something is not spiritually pure then it is spiritually impure. Hence, an error theory of judgments of spiritual (im)purity would be impossible. Clearly then, the formal objection to error theory overgeneralises. So while there is more to be said here I shall assume in what follows that error theory is not formally incoherent and error theorists can use the manoeuvre that I appeal to above to avoid to (B).

II.   *Why Everyone Must Be A Revisionist*

---

[11] See Christine Tiefensee, "Ought and Error", *Journal of Philosophy* 117: 96 114 (2020); Christine Tiefensee and Gregory Wheeler, "Why Formal Objections to the Error Theory Are Sound", *Analysis* (forthcoming); and Tiefensee and Wheeler, "Error, Consistency and Truth", *Nous* 56(3): 602-618 (2022).
[12] For example Bart Streumer and Daniel Wodak, "Why Formal Objections to the Error Theory Fail", *Analysis* 81(2): 254-262 (2021).
[13] See e.g. Perl and Schroeder 2019, and Eliot Salinger, "Defending the Kratzerian Presuppositional Error Theory" *Analysis* 81(4): 701-709 (2021).
[14] Streumer and Wodak 2021.

I have shown above how we can spare the worst of error theory's blushes. But by itself this is insufficient. Error theory still has counter-intuitive and troubling consequences for moral judgment: for example, that it is not true that the Holocaust was morally problematic. Unlike some error theorists I don't deny this. I simply claim that the competitor metaethical theories don't fare much better. They too have counter-intuitive and troubling moral consequences of their own. Indeed in some sense, theirs are worse. My argument for this surprising conclusion is based on a somewhat pessimistic reading of the last fifty years of moral philosophy; a recent history that shows if moral claims *can* be true, some of them are *bound* to be counter-intuitive and troubling. It's simply unavoidable. I provide two cases studies below: trolley problems and axiology. Each makes the same two points. Firstly, that it is impossible to satisfy all of our firmly held moral convictions and so, if we accept any moral claims, we will have to accept some counter-intuitive and troubling ones. Secondly, that error theorists alone can avoid this problem as they deny that we must accept *any* moral claims.[15]

II.i Trolley Problems

The trolley problems are a series of cases in which we are required to choose whether to bring about the death of a smaller number of people in order to avert the death of a larger number. In the classic case we choose whether to divert a lethal trolley away from a group of five people and toward a single person.[16] But should we, for example, push a fat man in front of the trolley to prevent it from killing a larger number? What about mining an innocent victim for organs in order to save the lives of five patients in need of transplants? Or what about turning a trolley on a track that runs in a loop where it will kill one victim and thereby not kill five? The task is to find a principle that gives the right result in all such cases.

Frances Kamm is author of much the most detailed work on the trolley problem.[17] Her *Principle of Permissible Harm* – though more detailed than I can do justice to here – states, in outline, that it is permissible to kill a smaller number in order to save a greater number if the act of killing is the very same act as the act of saving or is a 'noncausal flip-side' of it, but that it is not permissible to kill a smaller number in order to save a greater number if the killing is a mere means, in the causal sense, of the saving. This makes sense of intuitions in important cases. When you divert a trolley towards a smaller number, the act that kills them (the diverting) is the very same act as that which saves the larger number. But when you push the fat man, his death is not the same act as that which saves the larger number; it is a mere causal means to the saving of the larger number.

As I say, Kamm's actual articulation is more nuanced than this, but this will do to make the point we are interested in. It will do in part because it is roughly the articulation used by Shelly Kagan in his response piece to Kamm; a response that nicely brings out the unavoidability of troublingly

---

[15] A similar point could arguably be made using less up-to-the-minute sources. See, for example, Mackie's discussion of the upshot Sidgwick's 'dualism of practical reason' in JL Mackie, Sidgwick's pessimism", *The Philosophical Quarterly*, *26*(105): 317–327 (1974). The possibility that this might support an error theory is discussed in David Phillips, *Sidgwickian ethics* (New York: Oxford University Press 2011); and Christopher Cowie, "A New Argument for Moral Error Theory", *Nous* 56(2) 276-294 (2022).

[16] See e.g. Phillippa Foot, "The Problem of Abortion and the Doctrine of Double Effect", *Oxford Review* 5: 5-15 (1967) and JJ Killing, Letting Die and the Trolley Problem. *The Monist* 59(2): 204-217 (1976) for important early work.

[17] See Frances Kamm, *Morality, Mortality* (New York: Oxford University Press 1993); *Intricate Ethics: Rights, Responsibilities and Permissible Harm* (New York: Oxford University Press 2006); *The Trolley Problem Mysteries* (Oxford: Oxford University Press (2015).

counter-intuitive results in trolley cases.[18] Kagan is prepared to concede - for argument's sake if nothing else - that the principle does in fact better capture our intuitions about the relevant cases than any alternative he is aware of, but he is worried about the rationale behind it. Why, he asks, should we think that the distinction between being a mere causal means and being a non-causal flip-side is morally significant? This doesn't immediately *seem* like a morally significant distinction. It's not *intuitively* morally significant in the way that, say, the distinction between killing and letting die, or helping and not-harming is. Kamm claims that her distinction tracks a deeper distinction between *subordination* and *substitution*. Roughly, when we use someone's death as a mere causal means to saving others, we are subordinating that person to those who are being saved. This contrasts with cases in which the act that kills is the non-causal flip-side of the act that saves; in this case we are substituting the person killed for those saved. We subordinate the fat man, whereas we merely substitute the smaller number in the ordinary trolley case.

Does this respond to Kagan's worry? One might think so: the distinction between substitution and subordination seems a better candidate for being a morally significant distinction than that between being a cause and being a non-causal flip-side. There is, however, a problem. It is that, as Kagan sees it, Kamm effectively explicates the distinction between subordination and substitution in terms of the distinction between being a cause and being a non-causal flip-side. When explaining, for example, why a variant trolley case in which a smaller number are killed to save a larger number is a case of substitution rather than subordination – and so why the killing described in the case is permissible – Kamm claims that the killing of the smaller number is, in the case, merely the non-causal flip-side of the saving of the larger number. That, Kagan complains, won't do. The distinction between subordination and substitution is meant to serve as the rationale that makes sense of the moral significance of a distinction between causes and non-causal flip-sides. It won't do so if – as seems to be the case - explicating it relies on the very distinction it is meant to explicate.

Where does this leave us? Kagan suggests the following. Whatever principle best captures our intuitions about trolley cases – whether the Principle of Permissible Harm or some variant on it – that principle is going to tell us that certain properties or distinctions that we *don't* think are morally significant *are* morally significant; he asks us to imagine as illustration a moral principle that tells us that it is morally significant whether an act is performed on a Wednesday or a Sunday. That's roughly what's going on when the Principle of Permissible Harm tells us that the distinction between causes and non-causal flip-sides is morally significant. So we're left with a choice; a dilemma, really. One option is to accept a principle that captures our intuitions about trolley cases, but which comes at the cost of accepting the moral significance of a distinction that we really don't think is morally significant. The other option is to accept a principle that doesn't commit us to the moral significance of irrelevant distinctions, but that doesn't capture our intuitions about trolley problem cases. Both will involve substantial revisionism. As Kagan puts it, any genuine solution to the trolley problem will be one that is more reformist than we might initially have hoped.

To say that 'our final theory must be substantially more reformist than we might initially have hoped' is just to say that our final theory is bound to commit us to extremely counter intuitive claims. Kagan's preferred view is, of course, a form of consequentialism. This is a view that doesn't commit us to the moral significance of irrelevant distinctions, but certainly doesn't capture our intuitions about cases. On the contrary, it entails *famously* counter-intuitive and troubling moral results; perhaps, for example, that we should push the fat man or carve up the innocent victim for

---

[18] Shelly Kagan, "Solving the Trolley Problem" in F. Kamm (2015): 151-169.

his organs. So we face a dilemma. Put roughly, everyone will have to accept – something like - one of the following two claims:

> *Significance*: There is some distinction (e.g. causing vs being the non-causal flip-side), that we *really* don't think is morally significant, but that is.
> *Push & Carve*: We should (e.g.) push fat men in front of trains and carve up innocent people for their organs

Error theorists alone can avoid this dilemma (or some subtler articulation of it). Unlike everyone else they can deny *both* Significance *and* Push & Carve. They can do so because both Significance and Push & Carve entail that moral properties are sometimes instantiated. This is something that error theorists deny. So the trolley problem commits everyone but error theorists to one of these counterintuitive and troubling moral claims. Of course, error theorists will be committed to other counter-intuitive claims, but not these. I return to this below. But first, I present a second, rather different example that illustrates the same phenomenon: a commitment that error theory alone avoids to troubling moral truths.

II.ii Population Axiology

Derek Parfit's ground-breaking work in population axiology ended in failure. He could not find a theory that describes – to his satisfaction - the axiological properties of states of affairs in which the numbers of those affected vary from one state to another. Nevertheless, Parfit was optimistic that such a theory – Theory X – *could* be found.[19] Subsequent axiological investigations do not justify that optimism. On the contrary, they show that *any* axiology will be committed to extremely counter-intuitive consequences.

Consider for example Larry Temkin's 'spectrum argument'.[20] Imagine a relatively short, yet severe pain; being tortured for a day. Now imagine a pain that is only very marginally less severe but that lasts for many times the length of the first pain. Which is worse? Clearly, the second. Now imagine a third pain. This pain is only marginally less severe than the second but lasts many times its length. Surely, the third is worse than the second just as the second was worse than the first. More generally, it is surely the case that for any marginal decline in level of pain, there is some increase in its duration such that the lesser, longer pain is worse ('*The First Standard View*').[21] If this is true however then, provided that the 'better than' relation is transitive, there is a duration of the mildest possible pain (a mosquito bite, in Temkin's example) that is worse than the short, severe pain with which we began. This comparative claim is surely false. A day of torture is worse than a long life with a persistent mosquito bite. More generally, it is surely the case that for any sufficiently large decline in level of pain, there is no increase in its duration such that the lesser, longer pain is worse ('*The Second Standard View*'). Three extremely innocuous axiological claims – the First Standard View, the Second Standard View and Transitivity - are inconsistent. One of these claims must be false. Yet the denial of any one would be highly counter-intuitive and could lead to extremely troubling recommendations. Suppose, for example, that one were to decide that, all things considered, the best option is to deny the Second Standard View. One must then hold that for any sufficiently large decline in level of pain, there *is* some increase in its duration such that the lesser, longer pain is worse; it would be better to cause someone to suffer a persistently itchy mosquito bite than a day's torture. This is not only a strongly counter-intuitive result, but also deeply troubling as a recommendation. Call it the *Very-Non-Standard Second View*.

---

[19] Derek Parfit, *Reasons and Persons* (Oxford: Oxford University Press, 1984): 454.
[20] Larry Temkin, *Rethinking the Good* (New York: Oxford University Press, 2012): Chs. 2-6.
[21] This isn't exactly how Temkin phrases it, but it is close enough for our purposes.

The Very-Non-Standard Second View: For any sufficiently large decline in level of pain, there is some increase in its duration such that the lesser, longer pain is worse; it would be better to cause someone to suffer a persistently itchy mosquito bite than a day's torture.

The important point for our purposes is that this is a result that the error theorist – and only the error theorist - can avoid. Of course, the error theorist thinks that the Second Standard view is not true. This is for the simple reason that the error theorists thinks there are no moral properties.[22] But this does not mean that the error theorist is committed to the *Very-Non-Standard Second View*. On the contrary. The error theorist denies that there are any moral properties. And so the error theorist denies the *Very-Non-Standard Second View* too. The error theorist is therefore *not* committed to the view that it would be better to cause someone suffer a persistently itchy mosquito bite than a day's torture. So the error theorist – *and the error theorist alone* - avoids a strongly counter-intuitive and deeply troubling result when responding to Temkin's spectrum case.

Another closely related example comes from the work of Gustaf Arrhenius. Over the past twenty years Arrhenius has developed impossibility proofs that extend, improve and formalise Parfit's famous 'mere addition paradox'.[23] The proofs consist in a small number of axiological claims each of which is seemingly impossible to deny yet which, as a set, are provably inconsistent. Consider, as a particularly striking example, Arrhenius's sixth impossibility proof, in which the following propositions are shown to be inconsistent.[24] The first states pareto dominance of an unobjectionably weak variety:

The Egalitarian Dominance Condition: If population A is a perfectly equal population of the same size as population B, and every person in A has higher welfare than every person in B, then A is better than B, other things being equal.

---

[22] This holds on the assumption that moral error theory would, if true, take scope over axiological judgments too. I take it that it would. While error theorists do not deny that there are attributive goods (good cars, good banks, good websites), they do deny that there are predicative goods of the kind discussed by Parfit, Temkin and Arrhenius (below).

[23] See e.g. Gustaf Arrhenius, "An impossibility theorem for welfarist axiologies", *Economics and Philosophy*, *16*, 247–266 (2000); "The paradoxes of future generations and normative theory", in J. Ryberg & T. Tannsjo (Eds.), *The repugnant conclusion: Essays on population ethics* (Dordrecht: Kluwer Academic Publishers 2004); one more axiological impossibility theorem. In R. Sliwinski, L. Johansson, & J. Österberg (Eds.) *Logic, ethics, and all that Jazz, Uppsala philosophical studies*, (vol. *57*) (Uppsala: Department of Philosophy, Uppsala University 2009); The impossibility of a satisfactory population ethics. In H. Colonius & E. Dzhafarov (Eds.), *Descriptive and normative approaches to human behavior: Advanced series on mathematical psychology* (Singapore: World Scientific Publishing Company 2011); Population Ethics and Metaethics. *Iride: Filosofia e Discussione Pubblica* 25 (1):35-44 (2012); *Population ethics: The challenge of future generations* (Oxford: Oxford University Press, forthcoming). The view that Arrhenius's results could yield an error theory is explored in Cowie 2022, though that argument is perhaps better read as a reductio of the 'fixed points' view of Terence Cuneo and Russ Shafer-Landau, "The Moral Fixed-Points: New Directions for Moral Non-Naturalism", *Philosophical Studies* 171(3): 399-443 (2014) than as an argument for error theory. It is in any case reliant on reading the propositions that comprise Arrhenius's impossibility proof as conceptual truths; a reading that is not required by the present argument.

[24] Arrhenius forthcoming.

The second is a weakened version of the 'non-sadism' condition, according to which it is never better to add lives that are bad for the people living them than lives that are good for the people living them:

> *The Weak Non-Sadism Condition*: There is a negative welfare level and a number of lives at this level such that an addition of any number of people with positive welfare is at least as good as an addition of the lives with negative welfare, other things being equal.

The third condition states that a very marginal improvement in the welfare of a single person can, all else equal, be traded off for significant increases in welfare for a large enough number of other people:

> *The General Non-Extreme Priority Condition*: For any welfare level A and any population X, there is a number n of lives such that a population consisting of the X-lives, n lives with very high welfare, and one life with welfare A, is at least as good as a population consisting of the X-lives, n lives with very low positive welfare, and one life with welfare slightly above A, other things being equal.

The fourth condition effectively denies what is sometimes called 'the very repugnant conclusion': the view that negative welfare lives plus a large enough number of barely-worth-living lives is not worse than a small number of very good lives:

> *The Weak Quality Addition Condition*: For any population X, there is a perfectly equal population with very high positive welfare, and a very negative welfare level, and a number of lives at this level, such that the addition of the high welfare population to X is at least as good as the addition of any population consisting of the lives with negative welfare and any number of lives with very low positive welfare to X, other things being equal.

The final proposition is a much-weakened version of a principle that has elsewhere been called 'Non-Anti-Egalitarianism'.[25] Imagine a population with one person who is very well off and some other people whose lives are much worse. Now compare it to a new population of the same size as the first two combined, but in which everyone is only marginally less well-off than the best-off person in the first population. Which is better? Clearly, the new population:

> *The Non-Elitism Condition*: For any triplet of welfare levels A, B, and C, A slightly higher than B, and B higher than C, and for any one-life population A with welfare A, there is a population C with welfare C, and a population B of the same size as A∪C and with welfare B, such that for any population X consisting of lives with welfare ranging from C to A, B∪X is at least as good as A∪C∪X, other things being equal.

The propositions listed above are provably inconsistent (though I won't attempt to rehearse Arrhenius's technical proofs here). At least one must be false. Which? Suppose, for ease, that the first of them - *The Egalitarian Dominance Condition* – is false. This means that the following alternative must be true:

> *The Egalitarian Anti-Dominance Condition* If population A is a perfectly equal population of the same size as population B, and every person in A has higher welfare than every person in B, then A is *equally as good as or worse than* B, other things being equal.

---

[25] See e.g. Ng, YK, "What Should We Do About Future Generations? Impossibility of Parfit's Theory X", *Economics and Philosophy* 5: 135–253 (1989).

This is a *really* counter-intuitive claim. How could a population in which everyone has greater welfare be equally as good as, or even worse than an alternative in which everyone has lesser welfare? And it is not only strongly counter-intuitive. It will also yield deeply troubling recommendations; for example that one ought to distribute goods so as to ensure that everyone has lower rather than higher welfare levels.

Again, the important point from our perspective is that error theory avoids this troubling conclusion. Of course, error theorists will claim that *The Egalitarian Dominance Condition* – is not true. But they will not therefore have to claim that *The Egalitarian Anti-Dominance Condition* is true. On the contrary. Because *The Egalitarian Anti-Dominance Condition* entails that there are some instantiated moral properties, the error theorist will deny it. The error theorist is therefore *not* committed to the deeply troubling view that it would be better to make everyone worse off. So the error theorist – *and the error theorist alone* - avoids this strongly counter-intuitive and deeply troubling result.

III.   The Argument

The last fifty years of moral philosophy shows that if we allow that moral claims can be true, then among the truths we will find some extremely counter-intuitive and troubling ones. Their exact content will depend on who 'we' are but the claims may include the likes of, for example, *Push & Carve*, *The Very-Non-Standard Second View* or *The Egalitarian Anti-Dominance Condition*. Error theory alone avoids commitment to these. So it would be a mistake to think that error theory is the only metaethical position that lands its supporters with challenging moral commitments. It does not. Error theory has counter-intuitive and troubling commitments that the alternatives avoid, but the alternatives have counter-intuitive and troubling commitments that error theory avoids too. The balance is much more even than one might at first have thought.

But is the balance really even enough to allow us to say that error theory doesn't fare much worse than the alternative theories? Plausibly it is. This is in part because there is a case to be made for thinking error theory's *opponents* – i.e. realists, expressivists, and so on - actually face the *more* serious challenges. This is because the claims that they must accept are arguably *more* troubling and counter-intuitive. To see this, compare the worst consequences of each. Error theorists are committed to, for example:

(A) It is not true that the Holocaust was morally problematic.
(C) It is not true that it is wrong to discriminate against those with serious disabilities

The exact content of the claims to which error theory's opponents are committed will vary. But consider the following list:

- *Push & Carve*: We should (e.g.) push fat men in front of trains and carve up innocent people for their organs.
- *The Very Non-Standard Second View*: For any sufficiently large decline in level of pain, there *is* some increase in its duration such that the lesser, longer pain is worse.
- *The Egalitarian Anti-Dominance Condition* If population A is a perfectly equal population of the same size as population B, and every person in A has higher welfare than every person in B, then A is *equally as good as or worse than* B, other things being equal.

Which of these requires us to bite the bigger bullet? Is it worse to accept claims like (A) and (C), or to accept claims like Push & Carve, The Very-Non-Standard Second View and The Egalitarian

Anti-Dominance Condition? One may be inclined to say (A) and (C). They are more immediately arresting. It is however important to be careful here. Remember that error theory's commitment to (A) and (C) does not entail a commitment to:

(B) The Holocaust was morally unproblematic.
(D) It is ok to discriminate against those with serious disabilities.

This is important. While (A) and (C) are undoubtedly counter-intuitive and troubling, they are a lot less counter-intuitive and troubling than (B) and (D). Once we realise that error theory is not committed to (B) and (D) some of the sting is taken out of having to accept (A) and (C). By contrast, there is no equivalent manoeuvre by which non-error theorists can ameliorate their acceptance of Push & Carve, The Very-Non-Standard Second View, The Egalitarian Anti-Dominance Condition or similar principles. These principles don't merely state the absence of a moral property. They affirm its presence. And, in each of them, the conditions under which they affirm that property's presence is extremely counter-intuitive indeed. So there is a good case to be made for thinking that the worst consequences of error theory, morally judged, are actually less problematic than the worst consequences of the alternative metaethical views.

But aren't there nevertheless a great many *more* counter-intuitive consequences of error theory than of non-error-theoretic positions? And doesn't that mean the consequences of error theory for moral judgment are *in total* still much more problematic than those of the alternative metaethical views? This is an obvious response but it is too quick.

Firstly, it is far from clear how we are to individuate and enumerate the consequences of a theory. Indeed there is an obvious sense in which any of Push & Carve, The Very-Non-Standard Second View and The Egalitarian Anti-Dominance Condition – or indeed any moral principle - will have an *infinite* number of consequences; one for each possible specific state of affairs it could apply to. Even setting this to one side though, the reasoning that forces one to accept the troubling principle (i.e. Push & Carve, The Very-Non-Standard Second View and The Egalitarian Anti-Dominance Condition) will surely also force one to accept more, closely related troubling principles. Suppose, for example, that some non-error-theorist has been forced to accept The Very Non-Standard Second View. This principle is about *pain* but given the shape of Temkin's arguments, it will be difficult to avoid a series of closely related principles about harms other than pains, about pleasures, about benefits more generally, and so on. And so on. It is very difficult to quarantine. Troublingly counter-intuitive principles will inevitably proliferate.

Secondly, it isn't generally the case that a greater number of troubling consequences makes for consequences that are 'in total' more troubling. Quite generally, a theory with a large number of mildly counter-intuitive consequences will often fare better than an alternative with a small number of highly counter-intuitive consequences. Consider the following toy illustration. Suppose I am adjudicating between two theories of cheese. One entails that each of three artificial products that are borderline cases but on balance, you *do* think of as cheese - cottage cheese, burger cheese, cream cheese -, aren't cheese. A second theory has no such mildly counter-intuitive consequences, but does entail that the moon is cheese. The second theory has a smaller number of counterintuitive consequences about cheese. Clearly though, its consequences are 'in total' more troubling. Similarly with moral judgments. A theory with a large number of mildly counter-intuitive consequences will often fare better than an alternative with a small number of highly counter-intuitive consequences. And I have argued above that, plausibly, the worst consequences of non-error theoretic views are actually more problematic than those of error theory.

I should be careful not to oversell. My aim is not to show that the counter-intuitive consequences of error theory are in fact less problematic than those of the non-error-theoretic alternatives. My aim is much more modest. It is simply to show that matters are more closely balanced than error theory's opponents might at first think; error theory doesn't fare *that* much worse. The balance is close enough that, provided error theory does better on (e.g.) the metaphysics, epistemology and philosophy of language, it can and should be taken seriously as an option in metaethical space. I think that's hard to deny.

IV.   A More General Point

My defence of error theory has relied on two examples: the trolley problem and population axiology. Each illustrates the unavoidability of counter-intuitive and troubling moral claims. There are a number of reasons for choosing these two examples. Most obviously, one will matter to consequentialists (population axiology), one to nonconsequentialists (trolley problem). Another reason is that these are relatively well-known and easy-to-present. There are alternative examples that are perhaps less well-known or more technical - for example, some specific paradoxes and impossibility results concerning the nature of free choice and choice under risk –, or are rather more sprawling and difficult to present concisely; the non-identity problem and the procreational asymmetry arguably fit this description.[26] In these cases too, I think, accepting something really quite counter-intuitive will be unavoidable. There is an important general point here. Both the trolley problem and population axiology are *examples*. They exemplify the unavoidability of counter-intuitive and troubling moral results, but there are examples of the basic phenomenon everywhere we look. Indeed it is pretty much built into our workaday method of 'reflective equilibrium'; we must give up on some judgments to accommodate others in order to find a consistent balance. This is part and parcel of doing moral philosophy in the ordinary way. I have just listed some examples.

I shall venture one further speculation on this theme. The trolley problem and population axiology are areas of moral philosophy that have been studied in real detail; areas in which we have striven for precision and rigour and in which we have really subjected candidate answers to stress-testing, counter-exampling and so on. Indeed they are arguably the two areas of moral philosophy in the past forty years in which this method has been most conspicuously deployed. And interestingly, they are also two of the areas of moral philosophy in which the unavoidability of counter-intuitive and troubling moral results is probably most evident; everyone knows that the trolley problem seems to resist a happy solution and that population axiology yields paradox at every turn. So, although this is a small sample size - and there may be selection effects – there is something of a correlation between how hard and how rigorously we've looked in an area of moral philosophy, and the emergence of unavoidable, counter-intuitive and troubling moral results. And so I tentatively expect that we will find more and deeper counter-intuitive and troubling results the more we look in moral philosophy. This is something that I suspect won't come as a huge surprise to many working moral philosophers. Subjecting our existing moral judgments to scrutiny tends to highlight problems with them. All that I've really *added* to this

---

[26] See e.g. Johan Gustaffson's recent impossibility argument concerning the value of freedom of choice (Johan Gustaffson, "A Paradox for the Intrinsic Value of Freedom of Choice", *Nous* 54(4): 891- 913 (2019)); and Beckstead and Thomas's paradox concerning choice under uncertainty (Teruji Thomas and Nick Becksted, "A Paradox for Tiny Probabilities and Enormous Values", *Nous* (forthcoming). With respect to the non-identity problem and the ethics of future people, David Boonin's *The Non Identity Problem and the Ethics of Future People* (New York: Oxford University Press 2014) makes a persuasive case for the strongly counter-intuitive view that sub-optimal choices in non-identity cases are in fact permissible.

rather bland point is the observation that it comparatively bolsters error theory as against rival metaethical views.

Consider the following objections.

> *Objection 1.* "You claimed above that error theory's moral consequences aren't that much more counter-intuitive and troubling than those of competitor metaethical views, like realism. But that's wrong. The moral claims that a realist would have to reject in order to solve (e.g.) the trolley problem or the axiological paradoxes are a proper subset of the moral claims that error theorists reject. So there's an obvious sense in which error theory requires us to reject more moral claims than realism. Error theory is 'dominated'. And so there's an obvious sense in which error theory's moral consequences are significantly worse than those of realism."

When you're assessing how counter-intuitive and troubling realism's moral consequences are, don't just focus on what they must *reject*. Realists must also *accept* some moral claims that are highly counter-intuitive; for example, claims like Push & Carve, or The Very Non-Standard Second View or the Egalitarian Anti-Dominance Condition. Error theorists don't have to do this. And in fact, I claim, these troubling acceptances – acceptances that force themselves only on realists – are individually more troubling than any of the many individual rejections that follow from error theory. That helps to balance things out in terms of the overall appraisal of the moral consequences of error theory and realism respectively. Error theory's counterintuitive consequences are more numerous, though realism's worst consequences are plausibly worse.

> *Objection 2.* "You repeatedly say that metaethical views other than error theory have counter-intuitive and troubling moral consequences. That's misleading. Take realism, for example. It doesn't, *by itself*, have *any* moral consequences, as shown by the fact that it is perfectly consistent with *any* set of moral truths. This contrasts with error theory. Error theory's counter-intuitive and troubling consequences are a consequence of error theory alone."

I concede realism does not *by itself* entail any of the counter-intuitive claims that I have attempted to pin on it (e.g. Push & Carve, The Very-Non-Standard Second View and The Egalitarian Anti-Dominance Condition). Realists *could* avoid these or any other such claims, but the only obvious way to do so would be by denying that actions such as driving trolleys over victims have *any* moral properties, or that states of affairs that compare the distributions of pains or of welfare could stand in *any* evaluative relations to one another. This is an option in conceptual space but little more. If a realist ends up denying that riding trolleys over pedestrians is the kind of thing that has moral properties, then they have thrown out the baby with the bathwater. Why not just be an error theorist? That's a much cleaner way to arrive at the same result.

> *Objection 3.* "Suppose I grant you that any form of realism worth its salt will have counter intuitive consequences. I still don't think they are *troubling*, or at least not troubling *for realism*. Contemporary work in moral philosophy - like trolley problems and axiology –shows nothing more than that our moral intuitions are inconsistent. But that's not a problem *for realism*. It's a problem *for us*. Contrast this with error theory. Error theory isn't just inconsistent with *our* moral intuitions. It is inconsistent with *any* set of moral intuitions and indeed with *any* moral theory. That's a problem for error theory, not just for us."

There is a sense in which this objection succeeds, but it is not a sense that I am overly concerned with, at least not here. Distinguish between two reasons for rejecting error theory in light of its consequences. The first is that error theory conflicts with at least some of the moral claims that one *actually* holds or would actually hold if asked one's opinion; for example that the Holocaust was morally grotesque. I have shown that if *this* is one's reason for rejecting error theory, then one has a reason for rejecting realism too, because realism conflicts with moral claims we *actually* hold or would actually hold if asked too. In fact, realism conflicts more severely with some of these claims than error theory does. Now consider a second reason for rejecting error theory. It is that error theory conflicts with the true moral claims *whatever they may be* granted that some of them will look extremely counterintutive to us if presented with them at present. I concede that *this* reason for rejecting error theory doesn't generalise to a reason for rejecting realism. So I concede that my argument will lack suasive force against opponents of error theory who think in this way. But this does not concern me greatly. The common discomfort with error theory's consequences that I sketched in the introductory paragraph of this article comes much more from concern with its inconsistency with moral claims that people actually hold and would be uncomfortable giving up - like that the Holocaust was morally grotesque -, than from concern with its inconsistency with an unspecified set of claims, many of which we know we would find wildly counter-intuitive anyway. The former really makes people worry in a deep way about error theory. The latter does not. Or at least the worry that the latter expresses is very different – much less visceral and more and much theoretical and abstract – than the worry that I am addressing here.

> *Objection 4.* "This is a follow-up on the previous objection. I concede that *if* error theory counts as a view that we can take seriously, then your defence of it succeeds; we are all committed to some revision with respect to the specific moral judgments we make. Fine. But error theory isn't a view we can take seriously because of its *more general* moral consequences. It is beyond the pale because it entails that nothing is morally anything. That's just too much to countenance. It makes error theory a non-starter. And so your defence of it doesn't really do enough. We should still reject error theory because of its moral consequences."

There are two responses to this objection that work in tandem. The first is that I am presupposing that there are arguments *for* error theory in the background – familiar arguments from Mackie, Joyce, Olson, Streumer and others - though I am not providing those arguments here. They provide the basic reason to consider error theory. My aim is to defend error theory on the assumption that there is a basic case to be made for it already. The second response is that, as I noted in responding to the previous objection, the very general worry that error theory entails that nothing is morally anything, or that nothing is morally better or worse than anything else, is not, I think, the *real* worry. The real worry concerns error theory's more specific entailments; such as that it is not the case that the Holocaust was morally grotesque. And these are exactly the kinds of worries that my defence of error theory is set up to respond to; because, I have argued, when it comes to specific moral claims that we actually make, error theory and realism alike require us to make significant revisions. Part of the reason for thinking that the real worry is about error theory's specific - rather than very general - consequences is that, if we're honest, it is specific worries that come to the surface in ordinary discussion of the tenability of error theory. Suppose someone is thinking about metaethics and – based on broadly metaphysical concerns with, for example, irreducibly normative properties - comes to consider a revisionary view like error theory. You hope to convince them otherwise based on error theory's unseemly consequences. How will you do it? You might ask them: "Don't you think some things are better than others?", hoping that this will move them. But, as someone considering error theory already, it would not be at all surprising if your question were to fall flat. "No," they might answer "As an error theorist I don't think anything has moral properties, so of course I

don't think anything is better or worse than anything else!" The way to *really* pressure them into seeing why error theory's consequences are deeply problematic, would be to follow-up with something specific; something like: "But don't you think the Holocaust is worse than an act of selfless giving, or racism is worse than tolerance?". Answering "No" to this is *much* harder. The *real* challenge to error theory comes from these kinds of concerns. That's the challenge I've responded to.

> *Objection 5.* "Your above responses compare error theory to realism, and throughout your article when you speak of 'non-error-theoretic views' I get the impression that it is realists you are really addressing. But what about other 'anti-realist' views in metaethics, such as constructivism or expressivism? How does your argument work against these views?"

The fundamental challenge to error theory stands out clearly when compared with realism because realism, more than any other metaethical theory, appears well-placed to avoid commitment to troublingly counter-intuitive moral consequences. It stands out less clearly in comparison with other metaethical views because they have troublingly counter-intuitive moral consequences of their own. Consider, for example, constructivist views according to which what one ought to do is determined by one's desires. Views like these entail that the ideally rational Caligula should go ahead and run amok.[27] Error theorists needn't worry overly about comparisons with *these* metaethical views; at least not if the charge is that error theory entails troublingly counter-intuitive views. The pot needn't worry that the kettle calls it black. Nevertheless, there are versions of these anti-realist views that claim to fare better. There are, for example, forms of sophisticated 'objectivist' constructivism that claim to avoid the excesses of moral relativism, and there's modern expressivism in its quasi-realist guise.[28] These views are set-up precisely to avoid the troubling counter-intuitive moral consequences of their coarser predecessors.[29] They are set-up to ape realism's apparent moral neutrality. If they succeed in doing so, then their defenders will perhaps feel that they can claim the moral high ground over error theory. It is at this point that my defence of error theory kicks in. These sophisticated anti-realists, in virtue of aping realism's apparent moral neutrality, will be committed to whatever troublingly counter-intuitive moral claims result from ordinary moral philosophy: from the trolley problem, the puzzles of axiology (and so on).[30] So they cannot claim to hold the moral high ground after all. Or not that much higher at least.

---

[27] Classic discussions of these kinds of case are Gilbert Harman, "Moral Relativism Defended", *The Philosophical Review* 84(1): 3-22 (1975); and Bernard Williams, "Internal and External Reasons", in B. Williams, *Moral Luck* (Cambridge: Cambridge University Press 1981). More recently see Sharon Street, "In Defence of Future Tuesday Indifference", *Philosophical Issues* 19: 273-298 (2009), and for discussion David Sobel, "Subjectivism and Idealization", *Ethics* 119(2): 336-352 (2009); and David Sobel, *From Valuing to Value* (New York: Oxford University Press 2016): Ch. 13.

[28] Classic examples include Peter Railton, "Moral Realism" *The Philosophical Review* 95(2): 163-207 (1986); Michael Smith, *The Moral Problem* (Oxford: Blackwell 1994); Christine Korsgaard, *The Sources of Normativity* (Cambridge: Cambridge University Press 1996); Mark Schroeder Slaves of the Passions (New York: Oxford University Press 2007); Simon Blackburn, *Ruling Passions* (Oxford: Oxford University Press 1998); and Allan Gibbard *Thinking How to Live* (Cambridge, Mass.: Harvard University Press 2003).

[29] Though to call the former 'coarser' is perhaps unfair. They may be preferable if more 'sophisticated' views are guilty of the charge that David Enoch levels in his "Why Idealize?" *Ethics* 115(4); 759-787 (2005).

[30] The structure of this dilemma for sophisticated anti-realist views is familiar from Crispin Wright "Realism, Anti-Realism, Irrealism, Quasi-Realism", *Midwest Studies in Philosophy* 12(1): 25-49 (1988).

Objection 6. "I concede that if you think the task of moral philosophy is to articulate *general moral principles*, then counter intuitive and troubling moral consequences will be unavoidable. But that's a problem with moral generalism, not with moral realism. Particularists don't do moral philosophy in that way. And indeed part of the reason for this is precisely that (they think that) generalism leads to avoidable counter intuitive and troubling moral consequences.[31] So particularists are immune to your argument."

This is a good objection and my response is somewhat concessive. Particularists are better-placed than more traditional generalists to avoid the force of my argument *in some cases*. This is perhaps most obvious if we focus on the trolley problem. Here, the issue is in large part the search for a single general principle that can systematise our intuitions about a range of different cases in which we must choose between bringing about a harm to some and bringing about a harm to others. Particularists won't be engaged in this search. So they'll avoid the problem of trying to resolve it without counterintuitive and troubling moral commitment. But - even setting aside independent problems with particularism - there are limits to what this approach can achieve. To see this, let's contrast two claims that a particularist might make; one weak, one strong. According to the *weak* claim there are *some* areas of moral theorising in which one will be committed to counterintuitive and troubling results only if one falsely assumes there to be some general principle to be found. According to the *strong* claim *every* commitment to a counterintuitive and troubling moral result is the result of such an assumption. Particularists can certainly make the weaker claim. But it would be a very bold particularist indeed who would make the stronger claim. Yet unless they make the stronger claim, they too will have to admit some counterintuitive and troubling moral results and will to that extent vulnerable to my argument.

Why do I say that it would be bold for a particularist to make the stronger claim? While the trolley problem may be driven by the search - that particularists won't engage in - for a general principle, it is much less obvious *how many* difficult problems in moral philosophy have this structure.[32] In fact, they don't' seem to. Consider for example the axiological paradoxes. The particularist - same as anyone else - must still deny one or more of the propositions that comprise them. This is still very difficult. And isn't obviously made any easier by the fact that those propositions state relations between types of outcome rather than specific token outcomes. And, furthermore, the propositions in question have 'all else equal' clauses built into them; clauses that make them somewhat immune from much standard particularist counter-exampling (in which scenarios are dreamt up in which all else is not equal). So while there is some room for maneouvere for particularists, it is certainly not obvious that there is enough.[33] The particularist would need to show us that there is. To some extent this is something that needs to be worked through, case by case, for each of the various puzzles and paradoxes of moral philosophy. And so to some extent the success of my argument against particularists is as of yet undecided.

---

[31] A point made in Jonathan Dancy, "Ethical Particularism and Morally Relevant Properties", *Mind* 92(368): 530-547 (1983).

[32] See Andrew Sneddon, "Recipes for Moral Paradox", *American Philosophical Quarterly* 49(1): 43-54 (2012) for a taxonomy of moral paradoxes and subsequent argument that particularism is compatible with at least some.

[33] Perhaps the best option for particularists would be to reject general principles that underlie some of the paradoxes; for example, the transitivity of 'better than'. But now it is not obvious that it is really *particularism* that is doing the work as opposed to a more general attempt to work through the (axiological) paradoxes, looking for the most deniable principle.

Objection 7. "I don't buy your intended lesson from the trolley problem and the puzzles of population axiology. You make it sound as though we can't resolve these problems without taking on some highly troubling and counter-intuitive moral commitments. I think we can; seeing this just requires going into a bit more detail than you have done. With respect to the trolley problem, I don't find it *that* counter-intuitive or troubling that the distinction between causing and being the non-causal flip-side is morally significant. With respect to population axiology, both Temkin and Arrhenius's 'proofs' turn on claims that I am perfectly happy to reject. Temkin's turns on the transitivity of the 'all things considered better than' relation, which I am happy to reject given Temkin's own ground-breaking work in showing that goodness is essentially comparative.[34] Arrhenius's proof also turns on this, as well as on a view of welfare as modelled by the real numbers."[35]

This is a worrying objection. If there are solutions to the problems of moral philosophy (trolleys, population paradoxes and so on) that *don't* commit us to counter-intuitive claims, or if they do but the commitments aren't terribly concerning, then my defence of error theory fails. My response is that I simply do not think this is the case. While I have responses to each of the 'solutions' suggested above - about Kamm's principle, transitivity and the structure of welfare – I do not think it would be profitable to go into the details here.[36] That would be the wrong strategy. The basic response is simpler and more general. It is extremely rare to encounter a problem in moral philosophy for which there is a satisfying solution that does not require one to bite a fairly significant bullet. I would be deeply surprised to find a moral philosopher who honestly denies this. Yet it is just another way of saying that fairly strongly counter-intuitive claims are ubiquitous in moral philosophy.

Objection 8. "I found your presentation of the trolley problem and population axiology a bit misleading. When motivating the idea that, in each case, non error-theorists would have to take on a counter-intuitive commitment, you chose to work with the very worst-case scenario. In the case of the trolley problem, you imagined non error-theorists being forced to become caricatured utilitarians who permit carving up victims. But there are other options that look much less worrying. For example, they could just allow that the distinction between being a cause and being a non-causal flip-side is morally significant. That's counter-intuitive to a degree, but not very much. And in the case of Arrhenius's proofs, you imagined that non error-theorists are forced to abandon the Egalitarian Dominance Condition and accept the Egalitarian Anti-Dominance Condition instead. But again, this is a rather extreme

---

[34] Temkin 2011 Chs.12-13.

[35] Teruji Thomas, "Some Possibilities in Population Axiology", *Mind* 507: 807-832 (2018).

[36] With respect to Kamm's proof, while I agree that admitting the moral significance of a distinction between causing and being the non-causal flip-side is not *immediately* as troubling as, say, Push & Carve, it has knock-on consequences that are troubling in this way. This is discussed in detail in Tom Hurka, "Trolleys and Permissible Harm", in F. Kamm 2015: 135-151 in the context of his case of the killing of innocents via the shrapnel from bombing of a munitions factory. See also Otsuka's 'boulders' cases in response to an early formulation of Kamm's principle (Michael Otsuka, "Kamm on the Morality of Killing", *Ethics* 108(1): 197-207 (1997)). With respect to Temkin's work, Temkin agrees that the choice between abandoning one of either transitivity, the first standard view and the second standard view is unenviable and is at pains to point out that he does n*ot* take himself to have established the falsity of transitivity. With respect to Arrhenius's reliance on a certain view about the structure of welfare, the real issue is what the alternative is; Thomas's sketched alternative axiology – as I suspect he would admit – has troublingly counter-intuitive consequences of its own. Arrhenius (forthcoming Ch. 6) discusses the problems with accounts of this kind.

option. A non error-theorist would do much better to reject one of Arrhenius's *other* propositions. That would make their commitments considerably less counter-intuitive."

There is some truth in this charge. In order to make my argument stand out, I have perhaps saddled non-error-theorists with the worst-case moral commitments rather than the best-case. Nonetheless, this should not be over-sold. Firstly, while is true that I imagine non error-theorists taking on utilitarianism of an unspecified variety in response to the trolley problem, I was in fact – quite explicitly - taking my lead on this from Shelly Kagan; a philosopher who is hardly a straw-man and yet who defends a form of utilitarianism that recommends significant deviance from ordinary moral thought.[37] He is not alone in this. So in suggesting that the trolley problem may push us really quite a long way away from our ordinary, intuitive moral claims I am hardly trying to sell the reader a dud. Secondly, while I concede that the Egalitarian Dominance Condition is not the element of Arrhenius's proof that - if forced - I would give up on, I suspect that part of the reason for thinking that giving up on it would be the worst option is that it is the *simplest* principle in the proof. It is very easy to see what it means and so very easy to find it a strongly intuitive view. Other constituent propositions of his proof are, I think, comparably hard to deny; it is simply harder to see this. Consider, for example the Non-Sadism Condition. This is slightly harder to read but it effectively says – in fact, it says something weaker than - that it's better to add lives that are worth living than lives that aren't. So a non-error-theorist who chooses to give up on this will be committed to the view, which we might call *The Sadism Condition* that adding lives that are worth living is either equally as good as or worse than adding lives that are not worth living. This is *highly* counter-intuitive. Thirdly, even if some of the propositions that comprise the axiological impossibility results *really are* easier to deny than others, all are still such that their denial is counter-intuitive and troubling; in fact - given the realist's commitment to duality - *more* counter-intuitive and troubling than anything the error theorist is committed to. Consider for example *The Non-Elitism Condition*. This is arguably the easiest-to-deny of all of the propositions that comprise the impossibility result. Yet a realist who denied it would come very close to the view that, when comparing positive welfare populations, all that really matters is how well off the best-off person in either population is; even if they are only marginally better off than everyone else in the alternative population and the rest of *their* population is comparatively abject. That is an extremely uncomfortable result and any public policy that sought to realise it would rightly be criticised as one of the worst and most indefensible of policies in a long and inglorious history of policy-making. Fourthly, and more generally, my overall argumentative strategy succeeds to the extent that it *significantly narrows the gap* between the overall counter-intutiveness of the consequences of error theory and non-error-theoretic views respectively. Perhaps my presentation - of the trolley problem and population axiology – claimed a greater narrowing of this gap than I strictly ought. But it surely shows a significant narrowing of the gap nonetheless. Fifthly, even if one has specific responses in mind with respect to the trolley problem and population axiology, as I noted in section four, moral philosophy is riddled with puzzles and paradoxes. Everyone – or just about everyone – will run into results that they find very counter-intuitive and troubling somewhere.

> *Objection 9.* "I have some sympathy with your reading of the trolley problem and population axiology; I agree that from our present vantage point it doesn't look like there are many good options for solving these problems. But I'm nevertheless optimistic that some smarter, future people will be able to find some, even if it takes time. Isn't that a more moderate, sensible thing to think?"

Why the optimism? In moral philosophy quite generally, and particularly in axiology, the more detailed the work the more evident that troubling paradoxes and impossibility results become.

---

[37] Or compare Peter Singer, "Ethics and Intuitions", *Journal of Ethics* 9(3/4):331-352 (2005).

Objection 10. "There are many people against whom your defence of error theory won't work. Consider for example committed classical utilitarians who have very clear views about the right answer to all moral problems - trolley problems, problems of population axiology and so on – and don't find the consequences of these views troubling at all, at least not on reflection. These people will be troubled by the counter-intuitive consequences of error theory but not will not find the consequences of trolley problems or axiological puzzles – whatever they may be - at all troubling. So do you think that it is rational for these people to continue to reject error theory?"

I concede that my defence of error theory will lack *suasive force* against (e.g.) totally committed classical utilitarians who have internalised their view. However, whether it is *rational* for them to be unmoved in this way depends on whether it is rational to be unmoved by (e.g.) the fact that one's moral theory entails Push & Carve. I am not sure of the answer to this. Perhaps it depends on the epistemic context of the judger. I am inclined to think, however, that if we *do* allow that it could be rational to be unmoved by (e.g.) the fact that one's moral theory entails Push & Carve, then we should also allow that it could be rational to be unmoved by the counter-intuitive consequences of *error theory* too (e.g. that it is not true that the Holocaust was morally problematic). The consequences of each of these views are - to my eye at least – comparably distant from my considered moral judgment. Indeed, as I argued in the previous section, the former is arguably *further* than the latter.

VI.   Conclusion

Doing moral philosophy can feel like trying to complete a jigsaw in which the pieces don't fit together. The paradoxes, puzzles and impossibility results that moral philosophers have struggled with over recent decades really bring this to the fore. A commitment to counterintuitive and troubling moral claims seems unavoidable. This turns out to have far-reaching consequences for metaethics. Moral philosophy's various paradoxes, puzzles and impossibility results form the basis of a defence of its *enfant terrible*: error theory. The fundamental challenge that error theorists face is to explain how their view could be true given its inconsistency with hard-to-deny moral claims. Error theorists try to avoid this challenge. But they can take it at face value. It turns out that it is a challenge that *everyone* faces. Whatever metaethical theory you accept, you will have to accept some very troubling and counter-intuitive moral claims. Indeed realists may have to make some concessions that are *more* troubling than any that error theorists have to make.