A New Argument for Moral Error Theory

Moral error theory is the view that there is a systematic problem with moral judgment such that no moral judgments are true and anyone who makes one is thereby in error. Contemporary error theories trace the 'systematic problem' to the inescapable commitment of moral judgers to the existence of something that does not exist. There are two main subspecies.[i] According to the first (e.g. Joyce 2001) the non-existent troublemakers are categorical, normative reasons for action. According to the second (e.g. Olson 2014, Streumer 2017) the non-existent troublemakers are irreducibly normative relations. These arguments dominate the contemporary discussion of error theory. Structurally, they are very similar. My aim in this article is to present and defend an altogether different kind of argument for moral error theory; an argument based in the impossibility of reconciling inconsistencies among ordinary, first-order moral propositions. I refer to it as a first-order argument for error theory.

This argument does not figure in the contemporary literature but does have fairly recent historical precedent.[ii] Sidgwick famously identified three basic 'methods' of ethics: intuitionism, rational egoism and classical utilitarianism. Of these he regarded both rational egoism and classical utilitarianism as providing equally self-evident principles of practical reason. The result is the famous 'dualism of practical reason'. In the Concluding Chapter of the first edition of *The Methods of Ethics* Sidgwick describes this dualism as:

> "[A] contradiction so fundamental that if it cannot be overcome the whole system must fall to the ground..." (1874, 472)

Sidgwick, of course, does not use the expression 'error theory'. Nevertheless the view expressed above could from a contemporary vantage point be read very much in that spirit.[iii] Sidgwick is envisaging the whole system of practical reason "falling to the ground" and, as he later puts it, "…the Cosmos of Duty… reduced to Chaos" (Ibid., 473). The reason for this is not that moral properties (or, more broadly, principles of practical reason) are metaphysically suspect. It is rather that there is 'a contradiction' between competing first-order principles. My aim is to outline a contemporary argument in this spirit; an argument that moves from inconsistencies among first-order moral propositions to an error-theoretic conclusion.

Doing this requires defending two key claims. The first is that there is 'a contradiction' in first-order principles of the relevant kind. Sidgwick thought it lay in fundamental principles of practical reason. My focus will be narrower. I appeal to contemporary population axiology. Recent work in this field has shown that there is a very real threat of inconsistency among propositions that are very difficult to reject. The second key claim is that this contradiction entails error theory. This is my focus. I develop an argument based on the 'fixed-points' strategy in contemporary metaethics; a strategy according to which there are substantive moral propositions that have the status of conceptual truths. Putting these two claims together yields an error theory, or so I shall argue.

One reason for developing an argument along these lines is that it promises to have significant advantages over contemporary arguments for error theory. These advantages come primarily at the dialectical level. An argument based in first-order inconsistency promises to engage on ground shared by all, or at least many, parties to metaethical debate. In this respect it is unlike contemporary arguments. Consider the following five illustrations.

Firstly, a first-order error theory could, in principle, have traction against those contemporary moral philosophers of the emerging 'relaxed realist' tradition who would otherwise think themselves immune from error-theoretic worries (e.g. Kramer 2009, Dworkin 2011, Scanlon 2014). As these philosophers present it, contemporary error theories are misconceived because they are based on the mistaken assumption that we can coherently question - and reject - morality as a whole on *metaphysical* grounds ('external scepticism').[iv] The first-order argument for error theory, by contrast cannot be so easily rejected by relaxed realists. Their view is – as they explicitly acknowledge – compatible with a *morally* motivated questioning of morality itself. And so it is in principle at least compatible with the possibility of error theory based on inconsistency in first-order moral judgments ('internal scepticism').[v]

Secondly, the first-order error theory that I defend allows one to argue for error theory without committing on the big-picture metaphysical questions that divide many contemporary error theorists and their opponents. Consider for example Jonas Olson's argument for error theory. Olson's argument is based on denying that there are any irreducibly normative relations. This is, Olson claims, 'metaphysical bedrock' (2014, 136). It is an argument - as Olson freely admits – that will not have suasive force against those who do not already accept its conclusion. Olson and his opponents are now at something of an impasse. The first-order argument for error theory breaks this impasse. It doesn't rely on any high-brow metaphysical claims about the fundamental constituents of the universe. It just relies on the provable inconsistency of some ordinary moral judgments; something that one can accept whatever one's view on the nature of the metaphysical bedrock.

Thirdly, the first-order error-theoretic argument promises to speak to 'ordinary' or 'everyday' motivations for scepticism about morality in a way that contemporary metaphysical arguments do not. Consider for example Bart Streumer's recent defence of error theory. Streumer's argument is based on denying that there are any irreducibly normative properties. He argues for this via a principle of property identity; any two necessarily co-extensive properties, he claims, are in fact identical. Given this and given the supervenience of the normative of the descriptive, it follows that there are no irreducibly normative properties. An error theory (pretty much) follows.[vi] Streumer's defence of this argument is a model of good philosophy; it is thorough, rigorous and imaginative. But I am surely not alone in thinking that it is a justification of error theory that lies a long way from the kinds of considerations that initially move me – and many other philosophers and non-philosophers alike – to consider a broadly sceptical view of morality. Now of course arguments can be sound without accurately capturing one's initial suspicions - and 'initial suspicions' are probably not a very good guide to complex subjects anyway - but still, it would be *nice*, it would be *satisfying*, if there were an argument for the error theory that spoke, roughly, to them.[vii] An argument based in the impossibility of reconciling conflicting first-order moral propositions promises to fare rather better in this regard.

Fourthly, the first-order argument is well-placed to avoid the worrying over-generalisation of many contemporary metaphysical arguments for error theory. Consider the contemporary argument based on the denial of irreducibly normative relations (e.g. Olson, Streumer). It seems to entail more than a *moral* error theory. It seems to entail an error theory of judgements about epistemic normativity and prudential normativity too.[viii] This has proved off-putting to those who might otherwise be sympathetic; an error theory of epistemic normativity risks bringing with it nihilism about epistemic probabilities and even about beliefs.[ix] The first-order error theory needn't generalise in this way. Worrying inconsistencies in first-order moral propositions needn't entail analogous inconsistencies among first-order epistemic propositions.[x]

Fifthly, the first-order argument for error theory promises to have a wider appeal than the otherwise somewhat similar 'convergence-based' argument (e.g. Lillehammer 2004). This is an argument for error theory according to which if there were to be any true moral judgments, there would have to be convergence on their content amongst appropriately specified – e.g. rational, sincere and fully-informed – judges. Yet there is no such convergence. Hence, error theory.[xi] The convergence-based argument must draw a close connection between the moral judgment of appropriate judges and the moral truth. Many robust realists and traditional error theorists alike will simply reject this. The first-order argument for error theory fares better. It is – as I shall argue - perfectly compatible with a metaethic according to which moral truth is robustly independent of any judgment and in no way constituted or grounded by it.

So much for the dialectical appeal of the view. I shall now present the argument for it. I begin in section 1 by sketching the first-order inconsistency result from population axiology that forms the basis for my error-theoretic argument. In section 2 I present two unsuccessful ways of getting from here to an error theory. In section 3 I present a better route. In section 4 I reject some non-error-theoretic responses to my argument. The end result is a new argument for moral error theory that is worth taking seriously.

1. The Axiological Impossibility Theorems

It was an inconsistency between egoism and utilitarianism that led Sidgwick to worry that "the cosmos of duty would collapse to chaos". I shall work with a different – and, I think more worrying – example. It is taken from contemporary population axiology. I should stress that my concern is really with the route from such an inconsistency to error theory rather than with the details of the inconsistency itself. As such, my argument is somewhat conditional on the result from population axiology being as it appears: a genuine inconsistency among propositions that are extremely difficult to deny. I cannot claim to prove this here. To do so here would require a treatise in axiology that lies beyond the scope of this paper. I merely sketch the result. Nevertheless, I do take it to be genuinely troubling, much as it appears.

In his celebrated 'impossibility theorem' Kenneth Arrow (1950) provided a formal proof of the inconsistency of a small number of highly intuitive principles of social choice. His discovery revolutionised twentieth century economics.[xii] In recent years moral philosophers have begun to think seriously about whether similar results can be established in their own field; whether we can provide rigorous – perhaps formal - demonstrations of the inconsistency of a small number of comparably intuitive *moral* propositions. There is the very real prospect that we can. These theorems build largely on results first developed by Derek Parfit (1984). In the most famous of these – 'the mere-addition paradox' - Parfit argued that three very plausible axiological propositions entail a very implausible conclusion ('the repugnant conclusion'). The outline will doubtless be familiar to many readers.[xiii] So, however, will responses to it. The background assumptions used by Parfit are questionable and the conclusion is, on reflection, less counter-intuitive than it may at first appear.[xiv] Parfit merely set the ball rolling however. We can now reformulate his result using *much* weaker premises and background assumptions. We can also generate a much more objectionable conclusion. Furthermore, we can rigorously prove that the premises are inconsistent with the falsity of that objectionable conclusion.

Much of this owes to Gustaf Arrhenius. In a series of papers, and his ground-breaking work *Population Ethics: The Challenge of Future Generations*, Arrhenius outlines six 'axiological impossibility theorems'.[xv] In the sixth of these Arrhenius shows that the following propositions - some of which are recognisable weakenings of the premises of Parfit's original mere-addition paradox -

are inconsistent.[xvi] I provide the *very rough* gist of each as an orientation to the unfamiliar reader. Take some time to look at each. Aren't they all *obviously* true? Which would you deny?

The first proposition requires little by way of summary; it affirms a very weak form of pareto dominance for a fixed population size:

> *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.

The second proposition states a qualified and significantly weakened version of the claim that adding positive welfare lives to a population doesn't make it worse than adding negative welfare lives:

> *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 is more complicated but no less intuitive when grasped. With suitable qualifications it effectively denies a particularly extreme form of prioritarianism. According to this extreme version of prioritarianism a population of very good lives is worse than the same population size of significantly worse lives provided that there is one life that is marginally better at the otherwise inferior population.

> *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 states a qualified version of the claim that, all else being equal, adding very high welfare lives to a population doesn't make it worse than adding a mix of low welfare lives and negative welfare 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 an extremely weakened version of the principle that appears in Parfit's original paradox as 'Non-Anti-Egalitarianism'. The rough gist is that if everyone at one population is much better off than everyone at a second population, then adding *one person* to the second population who is *just marginally* better off than anyone at the first population doesn't thereby make the second better than the first. Though harder to grasp than the preceding, it is also no less intuitive when grasped:

> *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.

In what follows, I will take much of this at face value. I will assume both that the above propositions really are mutually inconsistent and that they are extremely difficult to deny. I will ask whether an error theory can then be deduced. Taking much of the axiological impossibility result at face value in this way is necessary; it is simply not feasible to provide a detailed treatment of it here. I will therefore not be engaging with the following two difficult, yet important issues in ordinary, first-order moral philosophy.

Firstly, I do not provide the technical proof of the inconsistency of the above propositions, nor do I detail the background assumptions about welfare on which it relies. Instead, I make a programmatic point that is more appropriate given my aims. My aim in this article is to sketch the shape of an error-theoretic argument based on first-order inconsistency. To this end the above axiological impossibility theorem is really a placeholder, albeit the best one that we currently have (perhaps have ever had). It is presented as an example of an apparent inconsistency amongst highly intuitive moral propositions. Anything that plays the same role equally well would do. I acknowledge that it is possible that it fails and that there is nothing else that plays this role. It is also a realistic possibility however that there are further, even better examples waiting to be discovered. We are all familiar with hard-to-resolve tensions in our own moral thought; tensions that emerge in the form of 'hard cases' when we try to balance competing values (quality vs quantity, freedom vs equality, and so on). Whether these familiar tensions rise to the level of irresolvable inconsistencies - as Sidgwick worried, Arrhenius's theorems suggest, and my error-theoretic argument requires – is yet to be discovered. We must put in the honest toil in first-order normative ethics to find out.

Secondly, I shall avoid taking a stance on big-picture questions in normative ethics with respect to the relation between the good, the right and virtue. Again, this is surely reasonable in the present context but I do acknowledge that some readers may find it unsatisfying with respect to my stated aim of establishing a *moral* error theory. I have in mind those who, while perhaps sympathetic toward the axiological impossibility result, do not take it to pose a problem - even in principle - for *morality* because (i) they do not take the theory of the good to play a significant role in the theory of the right or of virtue and (ii) they regard morality as concerned largely or primarily with the right or virtue. My immediate response to this is concessive. If someone wishes to argue that I 'only' end up with an axiological error theory (and not a properly *moral* error theory) so be it for now. That would still be a significant result. I do think however that, whatever one's preferred approach to moral theory, it will prove difficult to, as Parfit put it, 'quarantine' the axiological impossibility result. This is firstly because there has been significant recent work on translating the axiological paradoxes directly into deontic paradoxes - paradoxes that concern the wrongness of actions that bring about certain states of affairs rather than others - and secondly because even the most extreme non-consequentialist would surely assign her axiology a highly significant and indispensable place in her moral theory (it's just that she would not think that the axiological facts *entirely* determine the deontic or virtue-theoretic facts).[xvii]

2. The Route to Error Theory: Failed Arguments from Epistemic Security and Explanatory Grounds

We are trying to get from the axiological impossibility theorem to an error theory. How to do it? The axiological impossibility theorem aims to demonstrates an inconsistency in a set of moral propositions. Inconsistency is significant because it entails falsehood; given their inconsistency, we know that *at least one* of the propositions in the axiological impossibility theorem is false. To get from here to an error theory we would need to conclude that *all* moral judgments are

therefore false. This means that, at the big picture level, our argument must take the following form:

Master argument

1. If the propositions that constitute the impossibility proof aren't all true, then no moral propositions are true.
2. They aren't all true (they're inconsistent).
3. (1, 2) No moral propositions are true.

Add to this a further claim about the nature of moral judgment – which I won't defend here but which roughly goes under the label 'cognitivism' – and the error theory follows:

4. When we make moral judgments we're trying to state true moral propositions.
5. (3, 4) When we make moral judgments, we're thereby in error.

This captures the basic idea but leaves all of the hard work to be done. Why should one accept premise one of this argument? Why would the falsity of one of the propositions in the axiological impossibility theorem entail the falsity of all moral – or at least axiological – propositions? I can think of three ways in which we might go. The first is that the propositions in the impossibility proof have a particularly robust kind of *epistemic security*; they're the kinds of propositions that we're warranted in having as high a credence in as we are in *anything* (within the domain). The second is that the propositions in the impossibility proof are the *explanatory grounds* of other moral propositions; they purport to explain why other moral propositions are true. The third is that the propositions in the impossibility proof have the status of *conceptual truths*; they're the kinds of things that are made true, if at all, by the nature of the constituent moral concepts. I explore each of these options in turn. Each represents an improvement on the last.

2.1 The Argument from Epistemic Security

Consider a simple mathematical proposition: 1+1=2. Now suppose that a perfectly trustworthy and sane mathematical expert tells you that this is in fact false. Suppose that, given her status as a mathematical expert, you take her testimony seriously and come to believe that you do not know that 1+1=2. Surely your scepticism will not stop here. If you don't take yourself to know 1+1=2, what mathematical propositions *do* you take yourself to know? The falsity of 1+1=2 should *seriously* undermine your confidence in your basic mathematical belief-forming methods. You might reasonably conclude that you don't have *any* mathematical knowledge whatsoever. At least on a first pass, this would not be an unreasonable conclusion to reach.

Perhaps the propositions that constitute the axiological impossibility proof have a similar status; they have, with respect the moral domain, the same kind of seeming undeniability that 1+1=2 has with respect to the mathematical domain. They too are such that if you don't know them, your basic belief-forming methods are compromised and you don't have any knowledge in the relevant domain. And of course you don't know at least one of the propositions that constitute the axiological impossibility proof because they are inconsistent. So you don't have any moral knowledge. Call this the *epistemic security* strategy.

Could this help us to establish the first premise of our error-theoretic master argument? It is a promising start-point for two reasons. Firstly, it is at least somewhat plausible that a proposition like 1+1=2 has a kind of 'epistemic security' such that if one doesn't know it, then one doesn't have any mathematical knowledge. And secondly, the propositions in the axiological impossibility proof arguably do have the same kind of epistemic status with respect to axiology

as 1+1=2 does with respect to mathematics. Consider for example, the Egalitarian Dominance Condition:

> *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.

It is as hard, morally speaking, to see how this could be wrong as it is, mathematically speaking, to see how 1+1=2 could be wrong.

So far so good. Clearly though, this 'epistemic security' strategy will not do, no matter how it is fleshed out (as it would need to be). The problem is straightforward and instructive. The epistemic security promises to show how the falsity of one of the inconsistent axiological propositions generalises into a wider problem, but the wider problem that it generalises into concerns *knowledge*, not *truth*. It promises to show that we lack knowledge with respect to all moral propositions, not – as an error theory would require - that there are no truths among all of the moral propositions. For this reason the epistemic security strategy is the wrong kind of strategy to establish the error theory. The second strategy – from 'explanatory grounds' – promises to fare better.

2.2 The Argument from Explanatory Grounds

Both Sidgwick's dualism of practical reason and the Arrow paradox (mentioned at the beginning of section 1 as a model-of-sorts for the axiological impossibility theorems) threaten to demonstrate 'deep' or 'fundamental' problems with the principles of their respective domains; in Sidgwick's case the domain of practical reason, in Arrow's case the domain of preference-aggregation. This is in part because the assumptions about practical reason and preference-aggregation (respectively) on which they are based are good candidates for being *explanatorily basic*; they are assumptions that would explain why, if at all, other less basic propositions in the domain are true. An inconsistency in the *explanatorily basic* propositions will infect the whole domain.

Perhaps we can use this model to defend the first premise of our error-theoretic argument. Begin by supposing – not unreasonably – that on examining the commitments of ordinary moral thought we find it to have a certain kind of explanatory structure. At the base of the structure are some 'basic' moral propositions. These are the explanatory grounds of other less basic moral propositions.[xviii] The less basic moral propositions that they ground may in turn be the explanatory grounds – or part of the explanatory grounds - of less basic moral propositions still. Now suppose that it turns out that the propositions in the impossibility proof are good candidates for having the status of being explanatorily basic. The result would be a systematic error that infects the whole moral – or at least axiological - domain. And, unlike the epistemic security argument sketched earlier, the systematic problem isn't merely epistemological. It is of a sort that could ground an error theory and not just scepticism.

So this is a promising strategy for defending premise one of our master argument. Unfortunately though, it too fails. The first and most obvious reason for this is that relies on the claim that the propositions of the impossibility theorem turn out to be explanatorily basic in the moral domain. This isn't plausible. They lack the requisite generality and explanatory. To see this compare them to the traditional candidates for occupying the explanatorily basic role in the moral domain; contractualism, Kantianism, utilitarianism and so on (or to Sidgwick's own principles). These are very general propositions that could feasibly explain subsequent, more specific moral

propositions. The propositions of the impossibility theorem that we are considering are not like this. They are bad candidates for explanatory basicness in the moral domain.

Our prospective error theorist might try to dispute the picture of the structure of ordinary moral thought being presupposed above. She might claim that explanatorily basic propositions could be specific and plural. Perhaps, for example, Rossian deontology - in which the explanatorily basic propositions are a number of fairly specific pro tanto obligations - could serve as a model, or perhaps a broadly 'coherentist' as opposed to traditional 'foundationalist' structure of explanation better describes the commitments of ordinary moral thought. Suppose she is right. Let's just grant it. This would *still* not be enough for the error theorist. For suppose that, on analysing the explanatory structure of ordinary moral thought we establish that it is committed to the explanatorily basic status of these inconsistent axiological propositions. What conclusion could we legitimately draw from this? Surely, without some reason to think otherwise, the best we could do is to claim that ordinary moral thought turns out to be mistaken in its explanatory commitments. And from this, the appropriate conclusion would be a thorough-going revisionism of ordinary first-order moral judgment, not an error theory.

Again, this failure is instructive. To get to an error theory, the first-order inconsistency from which we are working must be an inescapable commitment of moralising as such and not merely a contingent but avoidable feature of the way that we happen to moralise. Any argument that we provide for error theory must speak to this. The explanatory grounds route – without supplementation - fails to do so. The third, conceptual route by contrast, promises to fare better.

3. The Conceptual Route

We have looked at two failed routes from the impossibility theorem to an error theory. Let's now look at a better route. It is to claim that the axiological impossibility theorem shows there to be a problem with our moral *concepts*. As it is an axiological impossibility theorem, the focus is on the concept of goodness. The obvious way to run the argument is as follows: The seeming undeniability of the propositions that constitute the impossibility theorem suggests that they are a kind of 'conceptual truth'; that their truth is explained or grounded by the essence of the concept of goodness. Now if the essence of a concept grounds a contradiction, then anyone who makes a judgment with that concept has failed to say something true (compare someone who makes judgments with the concept of a round square).[xix] This gives us the first premise of the master argument.

Compare this briefly with Sidgwick's dualism of practical reason, introduced at the beginning of the paper. Sidgwick's concern was based in part on the claim that the two competing principles of practical reason have the status of self-evidence. That is why the contradiction that they generate seems worryingly unavoidable. Now one way of understanding this is in terms of conceptual truth. So understood, the self-evidence of a proposition indicates the status of a conceptual truth; a self-evident proposition is the kind of proposition that is knowable to anyone with a mastery of the relevant concepts. If this is the case then a contradiction in self-evident propositions entails an error theory because it shows that the concepts used in practical reasoning ground a contradiction.

This *conceptual argument* has strengths of the two problematic arguments considered above without some of their weaknesses. Firstly, like both the epistemic security argument and the explanatory grounds argument, the conceptual argument promises to generalise from a problem with at least one moral proposition to a problem with *all* moral propositions. It does so because all moral judgments require moral concepts.[xx] Secondly, unlike the epistemic security argument, the

conceptual argument generates a *non-epistemic* problem; a problem that could yield an error theory and not merely scepticism. Thirdly, unlike the explanatory grounds argument the conceptual argument promises to show that there is a problem that is endemic to morality as opposed to being a merely contingent feature of ordinary moral thought. This is because the conceptual argument works by demonstrating an inconsistency in the concept of goodness itself. The result is an unavoidable, systematic problem with the axiological domain.

So the conceptual argument offered above promises to tick all of the boxes. But is it independently plausible? Could the propositions in Arrhenius's impossibility proof have the status of conceptual truths? Promisingly for our purposes, there is an existing, prominent view of the relationship between moral concepts and moral truth that plays pretty much exactly the role that we need in this respect. The view that I have in mind is Cuneo and Shafer-Landau's (2014) 'moral fixed-points' view. According to this view "there is a battery of substantive moral truths… that are also conceptual truths" (2014, 400). They give a list of examples:

- It is pro tanto wrong to engage in the recreational slaughter of a fellow person.
- It is pro tanto wrong to break a promise on which another is relying simply for convenience's sake.
- It is pro tanto wrong to humiliate others simply for pleasure.
- It is pro tanto wrong to torture others just because they have inconvenienced you.
- It is pro tanto wrong to impose severe burdens on others simply because of their physical appearance.
- It is morally admirable to express gratitude to a benefactor whose gift resulted from substantial sacrifice undertaken from exclusively altruistic motives.
- There is some moral reason to offer aid to those in distress, if such aid is very easily given and comes at very little expense.
- If acting justly is costless, then, ceteris paribus, one should act justly.
- The interests of others are sometimes morally weightier than our own.
- It is pro tanto wrong to satisfy a mild desire if this requires killing many innocent people.

We must now answer two questions. The first is whether substantial propositions like those identified by Cuneo and Shafer-Landau *are* conceptual truths. The second is whether, if they are, we would be warranted in putting the propositions from Arrhenius's axiological impossibility theorems on this list. Let's start with the second of these two questions. Here, I think, the answer is 'yes'. That's good news for our error-theoretic argument.

To see this it is useful to look at the four 'markers' that Cuneo and Shafer-Landau offer of conceptual truths as they are understanding them. The markers are offered as just that – markers – not necessary and sufficient conditions. The first marker is that if the proposition is true, it is necessarily true. The second is that the proposition is what they refer to as 'a framework proposition'. This is a proposition such that if you were to reject it, this would be good reason to think that you were not employing the concept in question. The third is that denial of the proposition would evoke bewilderment or bafflement amongst competent interlocutors. The fourth is that the proposition is a good candidate for being knowable simply by understanding the constituent concepts. Cuneo and Shafer-Landau's thought is that the examples listed above display these markers. If we look at the propositions that constitute the impossibility theorem, I think we see that they are equally good candidates for displaying these markers. Consider again, for example, one of the propositions in the impossibility theorem: the Egalitarian Non-Dominance Condition:

> *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.

This could fit perfectly on the example list above. It does just as well at meeting the four markers as those examples. This is *certainly* the case for the first and third markers. If it is true, it is necessarily true, and its denial would evoke bafflement for a competent interlocutor. The second and fourth markers ('framework proposition' and 'good candidate for knowability simply by understanding the concepts') are harder to assess but, on the face of it, *The Egalitarian Dominance Condition* looks just as good a candidate for meeting these markers as the propositions on Cuneo and Shafer-Landau's list. Admittedly, The Egalitarian Dominance Condition is a particularly clear example of this. But I think that the other conditions in Arrhenius's impossibility theorem listed above are also good candidates in this regard.

Consider an objection. One might think that there is an important difference in *complexity* between the kinds of propositions that figure in Cuneo and Shafer-Landau's list and in Arrenhius's impossibility theorem respectively. This comes out if we think about propositions other than the Egalitarian Dominance Condition. Consider, for example the Non-Elitism Condition. This condition begins with quantification over welfare levels and lives complex enough to make anyone scratch their head: "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…". This is *considerably* more complex than the conditions on Cuneo and Shafer-Landau's list. Doesn't this speak against including this kind of proposition on that list? I don't think it does. The relative lack of complexity in the propositions in Cuneo and Shafer-Landau's list is not deep. It is a result of a difference in parsing. We could, if we wanted to, for the sake of precision, equally well parse their propositions in (some approximation of) predicate logic in which case we would generate propositions no less intuitively complex than Arrhenius's. Consider, for example, the last proposition on their list: It is pro tanto wrong to satisfy a mild desire if this requires killing many innocent people. We could parse this more precisely as follows:

> For any person A, and any group of people B_1 to B_n, and any action φ that consists in the deliberate killing of B_1 to B_n by A, there is a strength of desire, L, such that it is pro tanto wrong for A to φ in order to satisfy a desire with strength less than or equal to L.

What we have done here is simply reparsed Cuneo and Shafer-Landau's condition on the model used by Arrhenius. The result is a proposition that is comparably 'intuitively complex' to the propositions that worried us in Arrhenius's theorem. So we shouldn't exclude Arrhenius's propositions from Cuneo and Shafer-Landau's list on the grounds of their being 'too complex'.

Putting these thoughts together, I think that we should conclude that Arrhenius's propositions are in fact very good candidates for going on Cuneo and Shafer-Landau's list. They meet the 'markers' just as well as the propositions on their original list, and they are not disqualified by the obvious objection. Of course this is not a result that Cuneo and Shafer-Landau will be happy with. They present their view as the under-pinning of an epistemology for moral realists. If I am right, however, it is the underpinning of an error theory.

The second, more interesting question facing my appropriation of the fixed points view is whether it is *correct* that there are substantial, conceptual truths of roughly the kind that Shafer-Landau and Cuneo identify. My argument for error theory requires that it is. In truth I am not sure of this, but I think it is plausible and I shall spend the remainder of this section defending it. Part of my optimism is based on the observation that *neither* of the prominent criticisms of the

fixed-points view in the contemporary literature – both of which I think are sound - actually undermine the aspects of the fixed-points view that are relevant *for our purposes*. They are both compatible with it. I have in mind two main lines of criticism.

The first line of criticism targets the role of the fixed points view in the overarching metaethic that Cuneo and Shafer-Landau are trying to establish. Their favoured metaethic is a robust form of nonnaturalist realism. The criticism is that the fixed-points view, contrary to what they claim, is incompatible with this.[xxi] This is a criticism that we can safely ignore; it is not relevant for our purposes. This is because the criticism does not target the fixed-points view as such. Rather, it targets the compatibility of the fixed-points view with a robust form of nonnaturalist realism. Now Cuneo and Shafer-Landau may be committed to a robust form of nonnaturalist realism but we are not (we are arguing for error theory). It doesn't matter for our purposes whether the fixed-points view is compatible with nonnaturalist realism or not. So we can safely ignore the first criticism of the use of the fixed-points view that we find in the literature.

The second line of criticism is similarly unproblematic for our purposes. It is that the fixed-points view promises to prove too much by illegitimately permitting one to draw existential inferences from the nature of concepts.[xxii] This criticism is brought out by thinking about the concept *God*. It is a conceptual truth about God that he is benevolent. So if the fixed-points view is correct, then we can infer from our competence with the concept *God* that God is benevolent. Clearly, though, this is incorrect; it entails that we can infer God's existence from the concept. We can at most infer that *if God exists*, then he is benevolent. Similarly, in the moral case. We can't infer from the concept *wrong* that torturing the innocent is pro tanto wrong. We can at most infer that *if anything is wrong*, then torturing the innocent is pro tanto wrong. So the fixed-points view fails.

For our purposes this objection poses no problem either. Suppose that the objection is sound. This shows that moral conceptual truths should be recast as conditionals with '*if anything is wrong/bad*' as the antecedent and substantive moral propositions as the consequents. This is perfectly compatible with the first-order argument for error theory that we are trying to run. That argument will now simply take the form of showing that the *consequents* of these conditionals are inconsistent. Given that they are inconsistent, the result will be that the antecedent - *if anything is wrong/bad* – will be false. From which we should conclude that nothing is wrong/bad. That is an error theory. So conditionalizing the fixed-points is perfectly compatible with the first-order argument for error theory.

So the two existing challenges to the fixed-points view in the literature are not a concern given our purposes. Are there challenges that *are* a concern? Perhaps the obvious concern is that the fixed-points view seems to entail implausibly that conceptually competent people will (thereby) recognise a whole range of substantive moral propositions as true. This is implausible because they don't; a lack of surety in moral judgment, not to mention disagreement, is evidence of this. Phrased a bit more precisely, the worry is:

1. For any moral proposition, M, and moral concept, C, which figures in M's content: if M's truth is explained by the essence of C, then anyone competent with C would thereby recognise M as true.
2. There is no *substantive* M and C figuring in M's content such that all judges who are competent with C would thereby recognise M as true.
3. (1, 2) There is no substantive M and C figuring in M's content such that the truth of M is explained by the essence of C.

This is a sensible objection. I am not convinced however, that it is sound. It is surely true that someone who defends the view that there are substantive conceptual truths is committed to saying *something* about what a competent possessor of the concept will believe. But as it stands the first premise of the argument just offered above is surely too strong. That the essence of a concept grounds a truth about it needn't entail that anyone who is competent with the concept will recognise that truth as a truth. There are all kinds of reasons that they may not that are consistent with their conceptual competence: they may, for example, be under the sway of a bad theory or an untrustworthy guru or both. Or perhaps there are other conceptual incompetencies or obstructive irrationalities in the background.[xxiii] Much more plausible is the weaker claim that such a person would be *in a position* to recognise the conceptual truth as a truth. So 1 above should be read as more like:

1*. For any moral proposition, M, and moral concept, C, which figures in M's content: if M's truth is explained by the essence of C, then anyone competent with C would thereby be in a position to recognise M as true.

With this point in mind we can make some headway toward blocking another one of the obvious objection to the fixed-points proposal. Someone might argue as follows:

> According to the fixed-points view the following is a candidate conceptual truth: it is pro tanto wrong to impose severe burdens on others simply because of their physical appearance. But this must be wrong. For suppose I were to meet some odious character who insisted on denying this proposition. A defender of the fixed-points view would be committed to saying that this person lacked conceptual competence. But we would clearly be better to say that they are immoral.

Bearing in mind that the fixed-points view is best expressed in terms of 1*, a defender of this view can actually agree with the above objection to a significant extent. They can say that the odious disputant may well *not* be conceptually incompetent. They may well be conceptually competent. What this means though is that the odious interlocutor is – in virtue of their grasp of the concept of wrongness - in a position to recognise that it is pro tanto wrong to impose severe burdens on others simply because of their physical appearance. This is consistent with this person's not actually recognising the proposition in question as true; there may be all kinds of distorting factor that prevent them from doing so. These are factors that may be best described as this person's being immoral.

All of this is encouraging with respect to my error-theoretic argument. So far the objections to the fixed-point view are either irrelevant for our purposes or based on a straw man version of it. There is however an altogether different kind of challenge to the fixed-points view worth considering; a challenge that is harder to respond to. It is that the arguments offered for thinking that some substantive moral propositions have the status of conceptual truths are *dialectically* weak; they are likely to do little to persuade those who are not already sympathetic. To see this it is useful to think about the 'markers' of fixed point status, as presented earlier. For each marker, working out whether a given proposition satisfies it requires prejudging whether that proposition is a conceptual truth or not. If one's priori judgment on this is negative, one will be unlikely to be swayed.

According to the first marker, if a fixed-point proposition is true, it is necessarily true. Now as it stands, this marker needs to be modified. Moral propositions may satisfy it (i.e. may, if true, be necessarily true) as a matter of *metaphysical* necessity or *normative* necessity rather than *conceptual* necessity. Clearly though, that would not establish that any moral truths are conceptual truths.

What the fixed-points theorist needs, then, is for the marker to read: if a fixed-point proposition is true, it is necessarily true *as a matter of conceptual necessity*. But how would one work out whether a proposition satisfies this marker without prejudging whether it is a conceptual truth? It is very hard to see. The second marker faces the same issue, even more obviously. It states that the fixed points are 'framework propositions' where this means that 'they are such that if you were to reject one, this would be good reason to think that you were not employing the concept in question'. Again, while this is surely a reasonable marker for a conceptual truth it is very difficult to see how one could apply it to a candidate proposition without prejudging whether that proposition is a conceptual truth. The third marker states that denial of the fixed-point proposition would evoke bewilderment or bafflement amongst competent interlocutors. Much like the first marker, this requires modification. There are many ways in which one can be bewildered or baffled. I may, for example, express my be bewilderment or bafflement when listening to a politician's inexplicable defence of his conduct, or when encountering a counter-intuitive discovery in fundamental physics. But these indicate nothing about violation of conceptual truths. In order for bewilderment or bafflement on hearing the denial of a proposition to serve as a marker for the fixed-point status of that proposition, the bafflement must be of the right sort. Specifically, it must be one that indicates a lack of conceptual competence. And whether the denial of a target proposition indicates *this* (i.e. conceptual incompetence) is impossible to ascertain prior to prejudging whether the proposition is a conceptual truth or not. The fourth marker (according to which the proposition is a good candidate for being knowable simply by understanding the constituent concepts), like the second, fairly clearly faces the same worry.

This is, to my mind, the major challenge to the fixed-points view, or at least to those aspects of it that relevant for our purposes. The arguments offered for it are dialectically weaker than one would wish; they are unlikely to persuade only those who are not already sympathetic to the conclusion. Of course, this doesn't mean that the conclusion isn't true or that the arguments for it are not sound. They may well be. It simply means that more needs to be done in order to advance the case for them persuasively in order to bring others on-board. From the perspective of my error-theoretic argument this is not a bad result; it is less a mark against my argument and more a direction for development.

4. Avoiding Error Theory?

The shape of my argument for error theory is now clear. The argument is doubly conditional. It is conditional on some substantive moral propositions being conceptual truths. It is also conditional on there being an inconsistency among some of those propositions. I don't claim to have conclusively demonstrated either conditional, but I do hope to have made the case that both should be taken seriously. Should the resulting error theory then be taken seriously as well? Or might there be a less radical alternative; a way of respecting what seems plausible about both the fixed-points view and the axiological impossibility theorem while avoiding the error-theoretic conclusion? I explain below why I am sceptical that there is a better alternative.

Perhaps the obvious option is to 'demote' one of the inconsistent set of axiological propositions from the status of fixed-point. For example, one might decide to demote the Egalitarian Dominance Condition. One would thereby arrive at a consistent *sub-set* of fixed-points consisting of all of those propositions that comprise the axiological impossibility theorem *minus* the Egalitarian Dominance Condition. An error theory would be avoided. The result would be counter-intuitive with respect to ordinary judgment, but much less so than error theory. It would therefore be justified. I'll call this the *basic demotion strategy*.

There are a number of problems with this strategy. A preliminary problem concerns *which* proposition to demote from the status of fixed-point. None are obviously better candidates than all of the others. I return to this below. The more serious problem is that it would not merely be necessary to demote a proposition *from the status of a fixed point*. It would also be necessary to deny *the truth of the proposition* itself. Only in this way would an inconsistency could be avoided. By demoting a proposition from the status of a fixed-point one would simply be creating the space to deny its truth without committing a conceptual error. And now the problem for the basic demotion strategy emerges. The propositions that constitute the axiological impossibility theorem are extremely difficult to deny. They are sufficiently hard to deny that, abandoning any one of them may, for a rational, sincere and conceptually competent person, simply not be an option. Think again, for example, of the Egalitarian Dominance Condition: of two perfectly equal populations, one in which everyone has a higher level of welfare is not, all else equal, worse. How *could* one deny this?

Now of course a sceptic may argue otherwise. 'It has not been *proven* to me', this sceptic may argue, 'that each of the propositions is so difficult to deny that it meets the conditions for being a fixed point'.[xxiv] My response to this sceptic is concessive. I cannot claim to have provided a proof that will satisfy a sceptic here. I can merely observe that I and many others see these propositions as obviously true, and as having satisfied key fixed-point markers, and therefore as good candidates for fixed points. In this sense, as noted earlier, my conclusion is necessarily conditional.

Consider another way of putting this point. I said above that demoting one of the propositions from the status of a fixed-point would be less counter-intuitive than accepting an error theory. Perhaps, however, this was too quick. Consider the following two options. The first option is that there are some axiological truths, and that there is no systematic problem with them, but that it is *false* that of two perfectly equal populations, one in which everyone has a higher level of welfare is, all else equal, not worse. This is the basic demotion strategy. The second option is that if there were any axiological truths, then one of them would be that of two perfectly equal populations, one in which everyone has a higher level of welfare is not, all else equal, worse, but there is in fact a systematic problem with axiology which means that there are no truths in that domain. This is the error theory. Is it really obvious that the latter is more counter-intuitive than the former? I am not sure (though again, I cannot claim to have a proof that would satisfy a sceptic o this point).

Consider, then, an alternative to the basic demotion strategy. It is a form of *pluralism*: there are many different yet internally consistent sub-sets of axiological propositions, none of which has a privileged claim to truth over any other. Each demotes a different axiological proposition from the status of fixed-point. One consistent sub-set would demote the Egalitarian Dominance Condition, another would demote Non-Anti Elitism, and so on. Each would be a different, equally acceptable, consistent set. This is appealing because it could help with the problems of the basic demotion strategy noted above. Firstly, it could help with the problem of deciding *which* proposition to demote from the status of a fixed point. Each proposition is demoted in some consistent sub-set, but none in all. Secondly, it could help with the problem of counter-intuitiveness of the basic demotion strategy sketched above; each proposition in the axiological impossibility theorem is true in all but one of the many consistent sub-sets and none need be false in all. Call this the *pluralistic demotion strategy*.

If anything, I think that this is actually a *worse* option than the basic demotion strategy. Let's just leave aside generic problems with moral pluralism. The pluralistic strategy is a bad one for reasons specific to the present context. This is brought out by contrasting it with a different

moral pluralism that one sympathetic to moral pluralism in general *might* reasonably find attractive. According to this view there are some moral propositions that any suitably reflective person would, on reflection, accept. These propositions form a basic framework from which to build a comprehensive moral theory but they under-determine its details; there are lots of different ways of accommodating them within a consistent and comprehensive theory. According to the somewhat attractive pluralist view, each comprehensive theory that consistently accommodates all of the basic framework propositions has an equally good claim to moral truth. This is the kind of moral pluralism that, as I say, one might reasonably find attractive if one were generally sympathetic to pluralism. Call it *reasonable pluralism*.[xxv] With this in place we can now see what is so problematic about the *pluralistic demotion strategy*. It is not appealing in this way. In reasonable pluralism, each fully comprehensive moral theory is consistent with the same common core of propositions that any reasonable and reflective person would accept. Contrast this with the pluralistic demotion strategy. Here, each of the many consistent sub-sets of axiological propositions will *deny* at least one of the highly intuitive propositions that constitute the axiological impossibility theorem (e.g. the Egalitarian Dominance Condition). That's not an attractive form of pluralism at all. Unlike reasonable pluralism, it is borne out of the inconsistency of a basic moral framework rather than the under-determination of a theory by a consistent framework.

The pluralistic demotion strategy does not improve on the basic demotion strategy. Consider then a final alternative: *choosing different normative concepts*. According to this view we should stop using the troublesome *concept* altogether – in this case the concept of goodness – and instead use a consistent alternative - call it goodness* – for which one has simply *stipulated* some set of mutually consistent propositions as conceptual truths.[xxvi] The result is *still* an error theory of attributions of goodness. It is supplemented however by a perfectly respectable theory of attributions of a new concept (goodness*) that, minus the attendant inconsistency, is otherwise identical. This is appealing. It promises to escape the core problem with the basic demotion strategy. The problem was that it is really hard to see how a rational, sincere and conceptually competent person could deny (e.g.) the Egalitarian Dominance Condition. By choosing to use a normative concept (goodness*) that, by stipulation, renders (e.g.) the Egalitarian Dominance Condition false, this problem is avoided.

Is this a satisfactory option? I don't think it is, although articulating why not is notoriously difficult. The obvious criticism is that the new concept, goodness*, is not a good concept to use. It is not a good concept to use because it entails that an obviously true claim (e.g.) the Egalitarian Dominance Condition is false. But this obvious criticism misses the mark. Goodness* does not entail that the Egalitarian Dominance Condition is false. It entails that *a modified version of* (e.g.) the Egalitarian Dominance Condition is false; a version stating that of two perfectly equal populations, one in which everyone has a higher level of welfare is not, all else equal, *worse**. Furthermore, even if it could be shown that goodness*, is not a good concept to use, it is not obvious what the relevance of this would be. After all, the new concept goodness* may still be a *good** concept to use (even if it is not a good one to use).

We have entered some extremely murky territory. It is difficult to clearly express what is wrong with the strategy of choosing different normative concepts. I do not think, however, that we have found any real reason to think that this strategy (*choosing different normative concepts)* resolves the problem with which we began. We have just found that articulating why is difficult. None of the three options canvassed for avoiding the error theory have been found to be appealing.

5. Conclusion

I have presented a new argument for the moral error theory, or, more modestly, a new version of an old argument. It is based on the provable inconsistency of a small number of moral propositions. I have merely sketched the argument here. There is a lot of work to be done at pretty much every juncture, but I hope to have shown that it is worth exploring.[xxvii]

Bibliography

Arrhenius, G. (2000). An Impossibility Theorem for Welfarist Axiologies. *Economics and Philosophy* 16: 247–266.

Arrhenius, G. (2004). The Paradoxes of Future Generations and Normative Theory. In J. Ryberg and T. Tannsjo (eds.) *The Repugnant Conclusion: Essays on Population Ethics*. Dordrecht: Kluwer Academic Publishers.

Arrhenius, G. (2009). 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.

Arrhenius, G. (2011). The Impossibility of a Satisfactory Population Ethics", in H. Colonius and E. Dzhafarov (eds.), *Descriptive and Normative Approaches to Human Behavior*: *Advanced Series on Mathematical Psychology*. Singapore: World Scientific Publishing Company.

Arrhenius, Gustaf. *Population Ethics*: *The Challenge of Future Generations*. Forthcoming with Oxford: Oxford University Press.

Arrow, K. (1950). A Difficulty in the Concept of Social Welfare. *Journal of Political Economy* 58: 328-346.

Barcan Marcus, R. (1980). Moral Dilemmas and Consistency. *The Journal of Philosophy* 77(3): 121-136.

Brink, D. (1988). Sidgwick's Dualism of Practical Reason. *Australasian Journal of Philosophy* 66(3): 291-307.

Chang, R. (2016). Parity, Imprecise Comparability and the Repugnant Conclusion. *Theoria* 82(2): 182-215.

Cowie, C. (2020). *Morality and Epistemic Judgment: The Argument from Analogy*. Oxford: Oxford University Press.

Cowie, C. (2020). *The Repugnant Conclusion: A Philosophical Inquiry*. Abingdon: Routledge.

Cuneo, T. and Shafer-Landau, R. (2014). The Moral Fixed Points: New Directions for Moral Nonnaturalism. *Philosophical Studies* 171(3): 339-443.

Diamond, C. (1996). Wittgenstein, Mathematics and Ethics. In H. Sluga and D. Stern (eds.) *The Cambridge Companion to Wittgenstein*.Cambridge: Cambridge University Press.

Donnellan, Keith. (1966). Reference and Definite Descriptions. *The Philosophical Review* 75: 281-304

Dworkin, R. (2011). *Justice for Hedgehogs*. Cambridge, Mass; Harvard University Press.

Enoch, D. (2011). *Taking Morality Seriously: A Defense of Robust Realism*. New York: Oxford University Press.

Enoch, D. (2014). How Principles Ground. *Oxford Studies in Metaethics*. New York: Oxford University Press.

Evers, D. and Streumer, B. (2016). Are the Moral Fixed Points Conceptual Truths?" *Journal of Ethics and Social Philosophy* 1: 1-9.

Fletcher, G. (2018). Pain for the Moral Error Theory? A New Companions-in-Guilt Argument. *Australasian Journal of Philosophy* 96(3): 474-482.

Foot, P. (1983). Moral Realism and Moral Dilemma. *The Journal of Philosophy* 80(7): 379-398.

Guttenplan, S. (1980). Moral Realism and Moral Dilemmas. *Proceedings of the Aristotelian Society* 80(1): 61-80.

Huemer, M. (2008). In Defence of Repugnance. *Mind* 117(468): 899-933.

Ingram, S. (2015). The Moral Fixed Points: Reply to Cuneo and Shafer-Landau. *Journal of Ethics and Social Philosophy*.

Joyce, R. (2001). *The Myth of Morality*. Cambridge: Cambridge University Press.
Joyce, R. (2019). Moral and Epistemic Normativity: The Guilty and the Innocent" in C. Cowie and R. Rowland (eds.) *Companions in Guilt Arguments in Metaethics*. Abingdon: Routledge.
Kalf, W. (2018). *Moral Error Theory*. Palgrave MacMillan: Basingstoke.
Killoren, D. (2016). Why Care about Moral Fixed Points? *Analytic Philosophy* 57(2): 165-173.
Kramer, M. (2009). *Moral Realism as a Moral Doctrine.* Oxford: Oxford University Press.
Kyriacou, C. (2015). Moral Fixed Points and Conceptual Deficiency: Reply to Ingram. *Journal of Ethics and Social Philosophy*.
Lillehammer, H. (2004). Moral Error Theory. *Proceedings of the Aristotelian Society* 104: 95-111.
Mackie, JL. (1974). Sidgwick's Pessimism. *The Philosophical Quarterly* 26(105): 317-327.
Mackie, JL. (1977). *Ethics: Inventing Right and Wrong*. Penguin: New York.
Mikhail, J. (2010). Rawls's Concept of Reflective Equilibrium and its Original Function in a Theory of Justice. *Washington University Jurisprudence Review* 3(1): 1-30.
Morton, J. (2020). Grounding the Normative: A Problem for Structured Non-Naturalism. *Philosophical Studies* 177(1): 173-196.
Parfit, D. (1984). *Reasons and Persons*. Oxford: Oxford University Press.
Phillips, D. (1998). Sidgwick, Dualism and Indeterminacy in Practical Reason. *History of Philosophy Quarterly* 15(1): 57-78.
Phillips, D. (2011). *Sidgwickian Ethics*. New York: Oxford University Press.
Olson, J. (2014). *Moral Error Theory: History; Critique; Defence*. Oxford: Oxford University Press.
Pettigrew, R. (2018). The Population Ethics of Belief: In Search of an Epistemic Theory X. *Nous* 52(2): 336-372.
Rowland, R. (2013). Moral Error Theory and the Argument from Epistemic Reasons. *Journal of Ethical and Social Philosophy* 7(1): 1-24.
Scanlon, T.M. (2014). *Being Realistic About Reasons*. New York: Oxford University Press.
Sen, A. and Maskin, E. (2014). *The Arrow Impossibility Theorem*. New York: Columbia University Press.
Shah, N. (2011). Can Reasons for Belief be Debunked? In A. Steglish-Petersen and A. Reisner (eds.) *Reasons for Belief*. Cambridge: Cambridge University Press.
Sidgwick, H. (1981). *The Methods of Ethics, 1st Edition*. Indianapolis, IN: Hackett.
Streumer, B. (2017). *Unbelievable Errors: An Error Theory About All Normative Judgments*. Oxford: Oxford University Press.
Temkin, L. (1987). Intransitivity and the Mere-Addition Paradox. *Philosophy and Public Affairs* 16(2): 138-187.
Thomas, T. (2018). Some Possibilities in Population Axiology. *Mind* 127: 807-832.
Vayrynen, P. (2013). Grounding and Normative Explanation" *Proceedings of the Aristotelian Society,* Supplementary Volume 87(1).
Williams, B. (1965). Ethical Consistency. *Proceedings of the Aristotelian Society*, Supplementary Volume 39, 103-124.
Williams, B. (1966). Consistency and Realism. *Proceedings of the Aristotelian Society*, Supplementary Volumes 11, 1-22.

[i] See Cowie (2014, Ch 1) for discussion. Olson (2014) claims that the second of the two sub-species is in fact the correct articulation of the first.

[ii] I focus on Sidgwick rather than the arguments developed in Williams (1965, 1966); papers in which Williams drew – or gestured toward - metaethical conclusions from moral dilemmas. This is because unlike Sidgwick, Williams' arguments - and the subsequent literature (e.g. Ruth Barcan

Marcus, 1980; Philippa Foot, 1983; Sam Guttenplan, 1980) are set up in a way that is somewhat orthogonal to my present interests and would require significant detour to do justice to. Firstly, Williams's argument and much of the subsequent literature, is framed as a case against traditional realist views of obligation and in favour of either a subject-dependent or non-cognitivist metaethic. I avoid engaging with these non-realist alternatives, focusing instead on the choice between whichever metaethical view is consistent with the propositions in the axiological impossibility theorems (which could in principle, be a range of different views, including sophisticated noncognitivism) and error theory. Secondly, Williams's argument is based in the inability of realist views to satisfactorily account for the 'residue' of reactive attitudes (such as guilt) when, in the context of a dilemmatic scenario, one fails to perform one of the required courses of action. The existence, and normative significance of, any such 'residue' is (legitimately) disputed and is not required in order to motivate the worrying inconsistency with which I am concerned here. Thirdly, much of the post-Williams literature has focused on the relationship between dilemmas, principles of deontic logic and inconsistency. I hope to bypass this as much as possible. This is in part because some of the disputed principles of deontic logic (most notably, the agglomeration principle) are rather specific to obligation, and not obviously applicable to the *axiological* inconsistencies with which I am concerned.

[iii] This is explored by David Phillips (2011), Ch.2. Phillips is picking up on Mackie's own reading of Sidgwick (Mackie, 1974).

[iv] Considering error theory, Dworkin writes: "What should I therefore do?... Grand metaphysical theories about what kinds of entities there are in the universe can have nothing to do with [it]… I reject Archimedean skepticism…the idea of an external, meta-ethical inspection of moral truth." (2011, 25).

[v] Dworkin writes: "I insist that any sensible moral skepticism must be *internal to* morality." (Ibid., italics mine).

[vi] The additional premise is required that moral judgments aren't true unless there are some irreducibly normative properties.

[vii] Compare David Enoch (2011 p. 8) on the motivation for his 'robust realism'.

[viii] Streumer claims his error theory entails an error theory of both epistemic and prudential judgment and that his error theory is therefore literally unbelievable, Olson claims that it entails neither. Guy Fletcher (2018) sides with Olson on the epistemic question but with Streumer on the prudential. It seems less likely that error theories based on categorical normative reasons entail epistemic (or prudential) error theories – a point made by Richard Joyce (2019) and Christopher Cowie (2019), though see Richard Rowland (2013) for the opposing view.

[ix] For the worry that it entails error theory about epistemic probabilities see Cowie (2019, Ch. 8). For the worry that it entails an error theory of belief attributions see Nishi Shah (2011) and Cowie (2019, Ch. 9). Streumer would claim the (literal) unbelievability of the error theory offsets some of the worst consequences of these.

[x] This is discussed in Richard Pettigrew (2018).

[xi] Although I contrast this with Streumer, one part of Streumer's (2018) error theoretic argument – the part in which he argues that moral properties are not identical with descriptive properties – actually makes a move rather like this. He claims that if normative properties were identical to descriptive properties, then under certain descriptive conditions judgers would be guaranteed to converge in their normative judgments, but in fact they do not.

[xii] For a summary of the theorem and its influence see Sen and Maskin (2014).

[xiii] The three plausible propositions and resulting implausible conclusion are, very roughly:
Mere-Addition: For any populations A and B: if A consists of lives at a positive, perfectly equal welfare-level, and B consists of the same lives plus any number of positive welfare lives, then, all else equal, B is not worse than A;
Non-Anti-Egalitarianism: For any populations B and C: if C has a higher total welfare, average welfare and is more equal, then, all else equal, C is not worse than B.

Transitivity: For any populations A, B and C: If B is not worse than A and C is not worse than B then C is not worse than A;
Repugnant Conclusion: For any population, A, that consists of lives at a high, positive, perfectly equal welfare-level, there is a population, Z, that is not worse than A, that consists of a larger number of lives at an only marginally positive, perfectly equal welfare-level.
For a nice summary based on slightly weaker assumptions than those provided below, see Huemer (2008).

[xiv] For critical discussion of the background assumptions see e.g. Temkin (1987), Chang (2016) and Thomas (2018). For discussion of the acceptability of the conclusion see Cowie (2020).

[xv] See for example Arrhenius (2000), (2011), (forthcoming).

[xvi] First presented in Arrhenius (2009).

[xvii] In (forthcoming Ch. 12) Arrhenius develops deontic versions. See also Arrhenius (2004).

[xviii] Though there are plenty of issues with respect to how exactly this works. See for example Enoch (2014) and Morton (2020). For a general orientation to grounding in the normative domain see Vayrynen (2013).

[xix] Perhaps this claim is less obvious than it may at first appear. Consider the following counter-argument to this, suggested by an anonymous reviewer. "Suppose that I am in a historical time and place where the concept of a king involves divine appointment: as a matter of conceptual necessity, an individual is a king iff that individual has been appointed as king by God. Now further suppose that there is some contradiction grounded in the concept of God. Then maybe you could defend an error theory about all claims about kings. Nevertheless it seems at least arguable that claims like "The king has imposed a new tax on imports" could still be true in such a context; for (you could argue) the phrase 'the king' may successfully refer to a real-world individual, who really has imposed a new tax on imports, even if the concept of a king is in some deep way conceptually broken." Although this counter-argument is intriguing, I do not think it succeeds. An utterance can succeed in picking out an individual within a context, even if what is said is false; a point familiar from the referential use of definite descriptions. The classic example is Donnellan's (1966) famous case of successful reference to a man drinking a glass of water, using the expression 'the man drinking the martini'. In the case of successful reference to the king, as described above, the false descriptive content that succeeds in referring is entailed by the concept itself (although there would remain a question with respect to the *mechanism* by which reference succeeds).

[xx] Though see Diamond (1996) for an alternative approach.

[xxi] See e.g. Killoren (2016).

[xxii] I am developing the argument from Evers and Streumer (2016). See also Ingram (2015).

[xxiii] Compare Kyriacou (2015).

[xxiv] This point was made to me by an anonymous reviewer.

[xxv] This reasonable pluralism should be recognisable as a modified-for-purpose version of a broadly Rawlsian 'pluralism based on over-lapping consensus'; a pluralism in which the considered first-order judgments with which reflective equilibrium begins spell out a broadly liberal view that is consistent with a range of competing comprehensive doctrines. I do not claim that this accurately captures any particular time-slice of Rawls's actual views on these matters, which are much disputed. For a useful discussion of some of the contested points of interpretation, see Mikhail (2010).

[xxvi] An alternative – suggested by an anonymous reviewer - is that we might *already* been using different concepts of goodness each of which is internally consistent. If this were the case then my argument for error theory would not get off the ground, as the fixed points would be fixed points *for different concepts*. Indeed, the argument would fail at an earlier stage than this as the axiological impossibility proof itself would be invalid. While this is (metaphysically) possible, (i) it would entail that some seemingly bona fide disagreements about the relative evaluative standing of evaluative states of affairs are not in fact disagreements, and (ii) I do not see what argument

there could be *for* it other than that it avoids the error theoretic conclusion for which I am arguing.

[xxvii] I am grateful to David Faraci, Lousie Hanson, Susan Notess, Chiara Brozzo, Bart Streumer, Christos Kyriacou and John Broome for comments and discussion. I am also grateful to the Universities of Stockholm and Cyprus for invitations to deliver this material.