

An Error-Theoretic Response to the Axiological Impossibility Results

Some things are obviously better or worse than others. It is better if good people are happy than if they are sad. It's worse if innocent people suffer a lot than a little. And so on. Suppose however we could show that a set of propositions, *each as obvious as these*, is inconsistent. This would be a troubling result. Yet it is exactly what axiological impossibility theorems show. They prove the inconsistency of some seemingly obvious propositions about what's better than what. The aim of this paper is to explore whether these theorems are of *meta-axiological* significance; whether they tell us something about the *metaphysical* nature of good and bad, and in particular whether they cast doubt on the very existence of facts about good and bad. I claim they do.

Consider an analogy. Suppose you are a traditional religious believer. There are some things that you think about the Deity and that you take to be fairly obviously true of it. One is that it created the Universe. Another is that it is wholly Good. You then reflect and become troubled. These two claims look worryingly inconsistent. How can they both be true given the existence of Evil in the world around us? Perhaps there is a good resolution. But it is worth taking seriously – as many have done – that the 'Problem of Evil' is the basis for questioning the very existence of a Deity.¹ Similarly with the axiological impossibility results. They can reasonably lead us to entertain meta-axiological suspicions about the nature – and even the existence – of good and bad.

So what is the meta-axiological view that I will be arguing for? Axiological *realists* think that there are some facts about good and bad ('axiological facts') that are 'stance-independent' – that is, not grounded in people's attitudes or beliefs or the social conventions they partake in – and that the judgments we make about them are sometimes successful attempts to represent those facts (i.e. beliefs).² *Anti-realists* deny this. Two variants of anti-realism concern us. One is *axiological expressivism*. This is the view that axiological judgments have the primary function of expressing attitudes, plans or some other conative state rather than representing the stance-independent axiological facts. My primary concern though is with *axiological error theory*. This is the view that

¹ A classic discussion of whether the Problem of Evil supports atheism is Plantinga 1974.

² I take the term 'stance-independence' from Shafer-Landau (2003) though my use differs slightly as Shafer-Landau's concern is morality generally, not axiology specifically.

axiological judgments are genuine attempts to describe the stance-independent axiological facts but that there are no such facts to be described. I shall argue that axiological impossibility theorems provide support for preferring error theory to realism. This is because the impossibility theorems present realists with the problem of having to accept some *highly* counter-intuitive axiological propositions; a problem that error theorists are uniquely well-placed to avoid.

The plan is as follows. In section 1 I provide some background to the axiological impossibility theorems and the meta-axiological suspicions that they arouse. In section 2 I present an arguments from the impossibility theorems to anti-realism that I won't be pursuing but that lays useful groundwork. In sections 3-4 I present and defend my own argument. Section considers an objection. Section 6 concludes.

1. Axiological Impossibility Theorems and Anti-Realism

In the closing section of *Reasons and Persons* Parfit famously expressed optimism that we will find the correct axiological theory. He referred to it as 'Theory X'. Concerning this theory he wrote:

Non-Religious Ethics is at a very early stage. We cannot yet predict whether, as in Mathematics, we will all reach agreement. Since we cannot know how Ethics will develop, it is not irrational to have high hopes.

1984, 454

Parfit's optimism has been an inspiration for many moral philosophers since - but is it warranted? *At the time of writing* perhaps it was. Parfit's work broke new ground. Optimism was not irrational. Now however the picture looks rather different. Forty years of hard graft have not led us to the land of axiological milk and honey. Quite the contrary. Contemporary axiological investigations have demonstrated worrying *impossibility results*; proofs of the inconsistency of a small number of axiological propositions each of which is extremely difficult to deny.

Consider for example Temkin's 'spectrum argument'.³ 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

³ Temkin 2012, Chs. 2-6.

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*).⁴ 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 mosquito bite. More generally, it is surely the case that for any sufficiently large decline in level of pain, there 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, Second Standard View and Transitivity – have proven to be inconsistent. Proofs like these undermine Parfitian optimism. They show – we might worry – that any theory of the good will entail the falsity of some axiological claim that any serious candidate for that theory must meet.

Another nice 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’.⁵ The proofs consist of 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.⁶ 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.

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:

⁴ This isn’t exactly how Temkin phrases it, but it is close enough for our purposes.

⁵ Arrhenius 2000, 2004, 2009, 2011, forthcoming.

⁶ Arrhenius forthcoming.

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 of 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 better 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’.⁷ 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

⁷ See e.g. Ng 1989.

AUC and with welfare B , such that for any population X consisting of lives with welfare ranging from C to A , BUX is at least as good as $AUCUX$, other things being equal.

Much as with Temkin's spectrum argument, I buy that the proposition in question are inconsistent and I find it very difficult indeed to see how I could deny any.

Some philosophers have begun to entertain the suspicion that these results are of *meta-axiological* significance; that they are best explained by some kind of axiological anti-realism such as expressivism or error theory.⁸ This suspicion is surely understandable. If investigation of some subject matter reveals a seemingly unavoidable inconsistency without a debunking explanation of that inconsistency, then it is surely reasonable to suspect that the problem may be with our presuppositions about the nature of the subject-matter itself. This is, I take it, the basic idea behind the use of the Problem of Evil as an argument for atheism. My aim is to explore whether these suspicions are borne out. I argue that they are. They support preferring error theory to realism.

1.1 What I Won't Be Doing

Before getting into the arguments I shall set the limits of my investigations. I do so by considering some sceptical voices I will be ignoring. Begin with the following three:

- i. "I don't know anything about these axiological impossibility results but I'm suspicious that they're as tough as you say. I'd need to study them in detail..."
- ii. "I don't buy the axiological impossibility results. Temkin's spectrum arguments are just sorites paradoxes and Arrhenius's impossibility results rely on modelling welfare distributions on the real numbers."
- iii. "I don't buy the axiological impossibility results. I'm a classical maximising utilitarian. I reject the anti-additive assumptions that underlie *all* axiological impossibility results."

These are common-enough reactions. I can't claim to respond fully to them here. To do so would require *doing axiology* whereas I want to *do meta-axiology*. So with respect to i and ii, I'm going to assume that either the axiological impossibility results really are troubling or that if they

⁸ The expressivist solution is discussed by Arrhenius 2012. Cowie 2020 defends the error theoretic view.

aren't there are some other results that are. Similarly with respect to iii; while classical maximising utilitarians may not buy Temkin's results or Arrhenius's, they famously do face their own axiological paradoxes.⁹ Of course I may be wrong about all of this. If so, the argument of this paper won't get off the ground. The argument of the paper is dependent on the axiological impossibility paradoxes being genuinely troubling.

Consider then the following rather different kinds of sceptical voice:

- iv. "I agree with you about the axiological impossibility results. They're very troubling. That doesn't surprise me greatly. I've always been suspicious of axiology. I'm not a consequentialist of any kind though, so it's no problem for me."
- v. "I find the axiological impossibility results troubling and I worry that maybe axiological anti-realism may indeed be true. But I don't take that to show that a broader meta-ethical anti-realism is true. The problem can be quarantined in axiology."

I am sceptical of these. I find it hard to see how an ethical theory can get by without an axiology. Furthermore, like Temkin and Arrhenius, I think that the basic structure of the impossibility results can be imported into the deontic idiom. And even if I didn't think this, I'd still suspect that there will be comparably worrying impossibility results for even the most committed of non-consequentialists. Suppose I'm wrong about all of this though. I would still be content enough to show that impossibility results yield interesting *meta-axiological* – if not *metaethical* – results.

Finally, there will be those who say:

- vi. "This is all old hat. People have worried for ages about whether moral dilemmas support anti-realism. Is there anything new here?"

As is often the case in philosophy, new debates can be re-reruns of older classics. There is undoubtedly literature in the background that I am failing to draw on that may be useful.¹⁰ I am

⁹ Their view entails 'The Repugnant Conclusion'. See Parfit 1984.

¹⁰ The classic literature on the metaethical significance of dilemma includes Williams 1965, 1966, Barcan Marcus 1980 and Guttenplan 1980. It is sometimes claimed that Sidgwick took axiological inconsistency ('the dualism of practical reason') to be of metaethical significance. See Phillips 1998, 2011.

confident however that the arguments presented and discussed in this paper are at least somewhat novel. And anyhow re-running old classics can be time well spent.

2. Just a Tool

How would an argument from axiological impossibility results to error theory go? My departure point is an interesting footnote on the final page of *Rethinking the Good*. Temkin quotes from correspondence with Shelly Kagan:

Various options in response to your arguments would be *much* easier to accept if we abandon realism in ethics (or never had it). After all, if morality etc is just a tool that we use to serve various purposes it might be that the best such tool is an inconsistent model! (If we aren't describing an independent realism, why think that the best tools are consistent?)

2012, 521 fn. 42

As I read it there are really two key claims here. One concerns realism. It is roughly:

Hard to accept: If realism is true, impossibility results force us to accept something that is hard to accept. That's a problem for realism.

The second concerns an alternative to realism. It is that:

Just a tool: If morality/axiology is 'just a tool', the impossibility results don't force us to accept anything so hard to accept. That's good for those views of morality/axiology.

Putting these two claims together we get the result that the impossibility results are easier to accept for views according to which morality is 'just a tool' than for realist views. Is this true? Is it a promising basis from which to claim that the impossibility results support anti-realism? We need to go slowly here.

Start with a simple point. Pre-reflectively, the axiological impossibility results mean that it will be necessary to accept at least one proposition that is very hard to accept. Think about Temkin's spectrum argument. Assuming the proof of the inconsistency is valid, realism implies the falsity of one of: The First Standard View, The Second Standard View, the transitivity of 'better/worse

than'. Accepting this, as Kagan rightly, notes, is hard. Or think about Arrhenius's sixth impossibility proof. Again, assuming the validity of the proof, realism entails one of the following is false: The Egalitarian Non-Dominance Condition, The Weak Non-Sadism Condition, the General Non-Extreme Priority Condition etc. Accepting the falsity of one of these would be hard. To say that something is 'hard to accept' is in the first instance to make a psychological claim. Things can be hard to accept for all sorts of very different reasons. It may be hard for me to accept the sudden death of a friend or family member because it is upsetting or unsettling. Some complicated equation may be hard for me to accept because I don't really understand it. These aren't the senses in which the axiological impossibility results are hard to accept. They are hard to accept because each of the propositions that comprise them seems obviously true and the negation of any one seems obviously false. Furthermore, there are no obvious available 'debunking' accounts; explanations of why we mistakenly regard a false proposition as obviously true. If a theory has 'hard to accept' consequences in this sense, it is the kind of thing that, generally, counts against that theory. Accepting the theory would require 'biting a bullet'.

What does any of this have to do with realism or the view of morality/axiology as 'just a tool'? Let's start with the idea that morality is 'just a tool'. The obvious way to flesh this out is in terms of meta-axiological expressivism: the view that the function of axiological judgment is to register approval or disapproval, convey one's plans or intentions, etc.¹¹ This is a view on which axiological judgment is distinctly (merely) 'tool-like'. Suppose it were true. It may then become an option that one could, *in principle*, make a series of axiological judgments that are mutually inconsistent *without going wrong in any way*. This is because one could – again, in principle – approve of all of, or plan with, an inconsistent set without thereby being in error. We can see this more clearly if we draw a contrast with realism. Realists think that the function of moral judgments is to represent truly. This function *cannot* be achieved whilst accepting all of an inconsistent set (given that there's bound to be some falsehood in there). So realists are committed to the existence of an error or mistake if one assents to all of the proposition in an impossibility theorem. Expressivists are not.

This summarises the very rough gist of the quotation from Kagan above. I will be making use of some of it when developing my own argument in the following section, more on which below. I won't however be making use of all of it. Specifically, I won't be making use of *just a tool*. Why?

¹¹ Classic discussions are Blackburn 1998, Gibbard 2003, Schroeder 2008.

As I read it, the idea behind *just a tool* is that expressivists would ‘find it easier’ than realists to accept inconsistencies and falsehoods. I’m broadly sceptical of how much load this can bear. One reason for scepticism is that inconsistent attitudes and plans don’t tend to make for good attitudes and plans. Someone who plans to go to the shop and also plans to not go to the shop is not someone who is planning well (unless their plan is to make unrealisable plans). Now of course there *may be* cases in which inconsistent plans are called for - Kagan’s claim, I note, is the modest one that ‘it *might* be that the best such tool is an inconsistent model’ (italics mine) - but if this is true of the impossibility results we’d need an argument for it. At present that’s not something we have. A second reason is that expressivists will surely want to be able to say that there is *something* problematic about *any* inconsistency. The reason for this is that inconsistency in plans or attitudes is the basis for expressivist explanation of behaviour of evaluative language in logical contexts. So, for example, traditional expressivists explain why instances of modus ponens with evaluative language are valid in terms of the inconsistency of the premises and the denial of the conclusion. Now presumably expressivists will also want to say that there is *something* wrong or problematic *in some way* (even if not *all-things-considered*) about invalid reasoning. Unless they allow that there is some in principle problem with inconsistent attitudes or plans, they won’t be able to do this.

I am sceptical, then, of arguing from the axiological impossibility results to anti-realism along the lines of Kagan’s quotation. I do think, however, that there is a better argument in this vicinity. The argument makes use of *hard to accept*. It also makes use of the big picture point that the axiological impossibility results are easier to accept if anti-realist views are correct. The difference is that the anti-realist view in question is *error theory*.

3. The New Argument for Error Theory (I)

Hard to accept tells us that the axiological impossibility theorems force realists to bite some bullets. In this section I show that the bullets that realists must bite are, in one important sense, significantly bigger than those that error theorists must bite. Focus on Arrhenius’s sixth impossibility theorem for now. It tells us that the following are mutually inconsistent:

- The Egalitarian Dominance Condition
- The Weak Non-Sadism Condition
- The General Non-Extreme Priority Condition

- The Weak Quality Addition Condition
- The Non-Elitism Condition

Suppose now that realism is true and you are trying to work out what to believe. You will have to reject one of the five propositions above; if not, you *know* you'll be going wrong in some way. None of them will be easy to reject; all, on examination, appear obviously true. But it may still be that some are easier to reject than others. Suppose, then, that on examination the proposition that looks easiest to reject turns out to be:

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.

You therefore reject this. So you are committed to:

(A) The Egalitarian Dominance Condition is not true.

This is common ground with error theorists. They also accept (A). The rationale is different though. Error theorists think that there are no axiological facts. They don't need to look at the axiological impossibility result to get to (A). They think:

(B) The Egalitarian Dominance Condition is not true because it falsely presupposes that the property of betterness is ever instantiated.

Realists by contrast think that the property of betterness *is* instantiated; just not under the conditions specified by the Egalitarian Dominance Condition. They think:

(C) The Egalitarian Dominance Condition is not true because it falsely describes the specific conditions under which something is better than something else.

These two competing rationales are very different; and they have different knock-on consequences. As we shall see below, (C) ends up committing the realist to some deeply implausible claims; claims that the error theorist avoids. To see this, let's go slowly, starting back with (A). (A) tells us:

(A) The Egalitarian Dominance Condition is not true.

Expanded, this means:

(D) It is not the case that: 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.

Realists and error theorists alike accept this. Presumably they also both therefore accept moving the negation that is taking wide-scope in (D), ‘in’ as follows:

(E) 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 not better* than B, other things being equal.

The move from (D) to (E) seems innocuous. It is an instance of the principle that: if it is not the case that A is better than B, then it is the case that A is not better than B. Again, both realists and error theorists will accept this. So far so similar. It is at *this point* that realists and error theorists diverge. The divergence is perhaps best seen by thinking about the following principle, which we can call ‘completeness’ and I state very roughly for now (I clarify it below):

Completeness: For any appropriate A and B: A is either better than B, worse than B, equally as good as B or on a par with B.¹²

This is a principle to which realists – but not error theorists – are committed. That error theorists are not committed to it should be obvious. Error theorists think that there are no axiological facts. That’s inconsistent with completeness. Realists on the other hand think that there are axiological properties. And so if A isn’t better than B, then – unless either A or B is not axiologically valenced at all – A must have some other axiological property; it must be worse or equal to/on a par with B. This has implications for what error theorists and realists respectively will infer from (E). Error theorists needn’t draw any inferences. But realists must think that if – as (E) states – population A is not better than population, then it must be either worse or equally

¹² ‘On a par’ follows Chang (c.g. 2002).

as good/on a par. The result is the following claim – rejected by error theorists but accepted by realists - which I will refer to as *The Anti-Pareto Principle*:

(F) 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 *either worse than, equally as good as or on a par with* B, other things being equal.

The realist is in a troubling position here. This is because (F), is very, very counter-intuitive indeed. It tells us (to simplify) that a population, A, that is better for everyone than another population, B, is nevertheless, all else equal, worse than it or equal to it. How could this *possibly* be? I find it very, very difficult indeed to see. So, in one sense at least, realism is clearly requiring us to bite a *much* bigger bullet than error theory. Realism is requiring us to accept the Anti-Pareto Principle. Error theory is not.

We can look at this point from several different angles. Firstly, we can look at it with a focus on error theory. Error theory is sometimes hung out to dry for its highly counter-intuitive consequences; it entails that all kinds of seemingly obvious moral and axiological propositions are not true. For example, it entails that it is not the case that Florence Nightingale was better than Hitler. This *is* counter-intuitive. Accepting it *does* require one to bite a bullet. I don't deny this. But its counter-intuitiveness can be overblown. It is easy to slide from the fact that error theory entails that it is not true that Florence Nightingale was better than Hitler to the view that error theory entails that they were equally good. But error theorists needn't slide in this way. They would do so only if they accepted completeness. But they don't. Their view is actually inconsistent with completeness.

Secondly, we can look at things with a focus on realism. Realism's traditional strength is that it delivers on our ordinary first-order moral and axiological judgment; it doesn't require us to revise those judgments in any way and so fares markedly better than error theory. The axiological impossibility theorems begin to call this traditional strength into question. The conjunction of the axiological impossibility theorems and realism entails that we would go wrong in some way unless we accept some wildly counter-intuitive first-order axiological claims: claims like the Anti-Pareto Principle. This is arguably more implausible than *anything* that the error theorist will ask us to accept. So realism's traditional strength over the error theorist is at least somewhat diminished.

Now of course all of this relies on the claim that realists must accept completeness. I haven't provided much of an argument for this. I think though that in a suitably hedged form, it is very difficult indeed for realists to reject. Suitably hedged how? There are pairs of propositions that perhaps realists need *not* think satisfy completeness. For example, it is not obvious that realists need think that the proposition 'Jupiter is larger than Saturn' stands in *any* axiological relation to the proposition 'the cat is on the mat'; arguably, the former is neither better than, worse than, nor equally as good as or on a par with the latter. That's because the states of affairs that these propositions describe aren't, as we might say, 'axiologically valenced' in any way. This means that the quantifiers in completeness shouldn't range over all states of affairs. But in order for the axiological impossibility theorems to create a worry for realists, we don't need them to. We need only worry about the propositions in the impossibility theorems themselves. Realists face a problem just in case each of *these* propositions – which concern the distribution of welfare over populations – satisfy completeness. And what reason is there for realists to think they do not? They are *clearly* not cases, like 'Jupiter is larger than Saturn' or 'the cat is on the mat' in which the states of affairs described are, as we might say, axiologically non-valenced.

We can actually strengthen the point. I have presented things in terms of completeness because it makes the contrast with the error theorist stand out clearly. But I am not actually convinced that the basic point I am trying to make – that realists will be committed to propositions like the Anti-Pareto Principle but error theorists will not - actually *needs* to be presented in this way. Completeness *may* turn out to be a red-herring. It depends on the details of the axiological impossibility proofs. The detail may show weaker or stronger things. The weaker option is that (e.g.) all but one of the propositions - The Weak Non-Sadism Condition; The General Non-Extreme Priority Condition; The Weak Quality Addition Condition; The Non-Elitism Condition – jointly entail the falsity of the other – The Egalitarian Dominance Condition. The stronger is that the first four propositions actually entail some specific proposition such as the *Anti-Pareto Principle*, which is then shown to be inconsistent with The Egalitarian Dominance Condition. The stronger entails the weaker, but not vice-versa. Now if the details of the proof take the stronger form, then there is in fact no need to appeal to completeness to show that realists must accept The Anti-Pareto Principle. The impossibility proof itself does all of the work. I do not want to rely on this, however, as it is rather more involved with the specific details of the axiological proof than I would like. My aim is to make a rather more general point about the relative

plausibilities of realism and error theory in light of the impossibility results. To this end, framing things in terms of completeness is probably a better way to go.

4. The New Argument for Error Theory (II)

Realists should not be worried yet. To this point I have made only a very limited claim. It is that with respect to *one* of the propositions in the impossibility theorems – the least plausible one, whatever it happens to be (e.g. Egalitarian Dominance Condition) - error theory requires us to bite a much smaller bullet than realism. That's good news for error theorists but it is far from sufficient to show that error theory is more plausible than realism; what about *all of the other* axiological propositions? What about The Weak Non-Sadism Condition, The General Non-Extreme Priority Condition and so on? Realists can simply say that they are all true, much as we would expect. Error theory by contrast will tell us *none* are true. Indeed this is the case for *all* axiological propositions, not just those that figure in impossibility theorems. Error theory will tell us that none of them are true. Realists will tell us that lots of them are true. So, overall, isn't it obvious that the first-order axiological consequences of error theory are much more counter-intuitive than those of realism? And so isn't it obvious that the error theorist must bite the bigger bullet? I'll call this *the numbers problem* for error theory.

The answer may end up being 'yes' but in this section I explain why we should be cautious. The claims of the previous section may be much more far-reaching in their significance than we have seen to this point. In short, the numbers problem may well be over-stated. In 4.1 I argue for this by making some general remarks on how we should – and should not – aggregate the counter-intuitive consequences of a theory. In 4.2 I argue that the number of counter-intuitive consequences of error theory and realism respectively may be more closely aligned than one would at first expect. In 4.3 I put these thoughts together with the argument of section 3 to make a case for error theory over realism.

4.1 Aggregating Number vs. Degree

Many theories, especially in philosophy, will have counter-intuitive consequences. We will often find ourselves having to adjudicate between these theories. There are a number of different factors that we will have to weigh up. One is the *number* of counter-intuitive consequences. Another is their *degree* of counter-intuitiveness. There are many more factors, doubtless, but

let's just start with these two. How should we balance them? Naïve renderings of the numbers problem can tacitly presuppose that 'the numbers count' in an implausibly strong way; a way in which numbers *swamp* degree. To see this, consider a very simple additive model according to which 'numbers are trumps':

First Model: For any two theories T_1 and T_2 of some subject-matter, S , if T_1 has more counter-intuitive consequences than T_2 , then – with respect to their counter-intuitive consequences – T_1 is less plausible than T_2 .

Clearly, this is a bad model. Suppose I am toying with two different theories of what it is for something to be cheese. One theory has the result that many artificial products that you might ordinarily think of as cheese and that we ordinarily call cheese (e.g. cottage cheese, burger cheese) aren't cheese. A second theory has no such mildly counter-intuitive consequences, but does entail that the moon is cheese. This demonstrates the obvious falsity of the first model; sometimes, the degree of implausibility of some consequence can trump a great number of less implausible consequences. This shouldn't come as a big surprise. Hopefully, it also provides some direction as to how an error theorist might begin to respond to the numbers problem. Naïve renderings of the numbers problem are premised on aggregative principles that resemble the first model. Error theorists should reject these in favour of an alternative model avoids swamping. Consider then a second model. According to this model, more numerous counter-intuitive consequences is only worse *all else being equal*:

Second Model: For any two theories T_1 and T_2 of some subject-matter, S , if T_1 has more counter-intuitive consequences than T_2 , then *all else being equal* – with respect to their counter-intuitive consequences only – T_1 is less plausible than T_2 .

This model is better. The 'all else equal' clause leaves the space to avoid swamping. Although a theory of cheese that entails burger cheese and cottage cheese are not cheeses has more counter-intuitive consequences than a theory on which the moon is cheese, all else *isn't* equal; the counter-intuitive consequence of the latter are wildly *more* counter-intuitive than any of those of the former. For the very same reason though, the second model doesn't – without significant further work – yield a version of the numbers problem that should bother an error theorist. All else *is not* equal between error theory and realism; while error theory has a greater number of counter-intuitive consequences, the most counter-intuitive of those consequences (e.g. the Anti-

Pareto Principle) is much more counter-intuitive is, I submit, more counter-intuitive by some way than any of the counter-intuitive consequences of error theory.

I shall not attempt to find the right aggregative model (I'm pretty sure I couldn't). I don't need to. It will suffice to claim – as is surely correct – that the right aggregative model will have to strike a balance between the number of counter-intuitive consequences and the degree of their counter-intuitiveness. Simply being aware of this can go some way to offsetting the numbers problem and to levelling the playing field between the error theorist and the realist; at least when it comes to evaluating the overall counter-intuitiveness of their views.

4.2 The Number of Counter-Intuitive Consequences

I have simply conceded to this point that error theory will have *a greater number of* counter-intuitive consequences than realism. Indeed for all I have said to this point, it may look as though realism will have only as many counter-intuitive consequences as there are axiological impossibility theorems - perhaps just one or two – whereas error theory will have *a huge number* of counter-intuitive consequences. This characterisation is mistaken. The number of counter-intuitive consequences of error theory and realism respectively is, I think, likely to be much more even than the initial presentation of the numbers problem suggests.

Consider Temkin's 'spectrum' impossibility theorem again. Suppose we are presented with a specific version of it and suppose that we decide that the proposition that is the least plausible is:

- (G) For some very painful pain of one day in length and some second, only ever so slightly painful pain, there is no duration of the latter pain such that experiencing it is worse than experiencing the former.

So, if we're realists, we'll reject (J). Presumably though, we won't be able to stop here. We will also have to reject more propositions that are very similar to (J), such as:

- (H) For some very painful pain of *thirty six hours* in length and some second, only ever so slightly painful pain, there is no duration of the latter pain such that experiencing it is worse than experiencing the former.

- (I) For some very painful pain of *two* days in length and some second, only ever so slightly painful pain, there is no duration of the latter pain such that experiencing it is worse than experiencing the former.

Indeed we will surely have to accept an infinity of claims of this type. And it's not just that we're having to reject many types of the same token either. We would *also* have to reject similar claims about other types of harms, not just pains, given that the impossibility result can be constructed for these too. So, we will have to reject:

- (J) For some *very bad but non-painful harm* of a day in length and some second, only ever so slightly less bad such harm, there is no duration of the latter harm such that being subject to it is worse than being subject to the former.

With this, as with (K) and (L), we will presumably have to reject an infinite number of tokens of type (M) too, such as:

- (K) For some very bad but non-painful harm of *two days* in length and some second, only ever so slightly less bad harm, there is no duration of the latter harm such that being subject to it is worse than being subject to the former.
- (L) For some very bad harm of *thirty six hours* in length and some second, only ever so slightly less bad harm, there is no duration of the latter harm such that being subject to it is worse than being subject to the former.

Nor does it stop here. We will also have to reject (sort of) contrapositive claims about *benefits* or positive experiences as opposed to harms too, given that the basic structure of the impossibility results work equally well with these. This means rejecting:

- (M) For some very great positive experience of a day in length and some second, only ever so slightly less good positive experience, there is no duration of the latter benefit such that being subject to it is better than being subject to the former.

I hope it is easy enough to see where I'm going with this. Realists are going to have to reject *a lot* of *very* plausible axiological claims. And for each one, recall, they're not just going to reject it in favour of its negation (as an error theorist might); they're going to have to reject it in favour of a

highly implausible, very hard to believe alternative axiological claim with it is inconsistent – a claim like the *Anti-Pareto Principle*. It would be quite wrong, then, to think that realists will have to accept *only a few* counter-intuitive propositions, in contrast to the error theorist's acceptance of a great many. Both will have to accept a huge number, perhaps an infinity.

Now this might strike some people as too quick. As I have presented it, each of (G) through (M) is a *different* proposition that the realist will have to reject in favour of some alternative, implausible axiological proposition. But is this right? Won't it be possible to find some *single, general* principle that covers all of (e.g.) (G) through (M) such that realists must reject only this one, single principle (and all entailments of it)? And by contrast, won't error theorists have to reject a *great many* general, plausible principles? This is a sensible worry but I don't think it actually demonstrates what it is meant to; a significant difference between the numbers of counter-intuitive consequences that error theorists and realists respectively will have to accept.

Return to Temkin's axiological impossibility results. These are results that have a general shape. The shape concerns an inconsistency between a proposition about how quantity can (axiologically) trump quality, a proposition about how quality can axiologically trump quantity, and some kind of axiological transitivity proposition. It is very hard to know how finely grained or coarsely grained we should go when working out how many propositions these results involve (and so, how many propositions we will have to end up rejecting, whether as realists or error theorists). There will be very finely grained views according to which the number of propositions is very large. These views will include, for example those of *particularists* who claim there to be an infinite number; one corresponding to each token. Close-by will be those – call them *∀-phobes* after a former teacher of mine who once declared himself 'allergic to universal quantification' – who are generally disinclined to high-levels of generality in moral explanation. This kind of view could be argued for on the basis of any number of considerations, ranging from pessimistic induction to the view that general laws are rightly regarded with scepticism in the social sciences, and so one might reasonably expect moral laws to be roughly similar. At the other extreme are those inclined toward maximal generality. These *maximal generalists* – think Kant or Parfit – are not shy of universal quantifiers (and boxes) and will likely think it possible to identify a small number of principles at a very high level of generality that cover all instances of impossibility results of a given shape. Between these two extremes there will of course be moderate views, covering a wide range.

My point is this. There are different approaches to the identification of moral or axiological principles; some yield fine-grain, some coarse-grain. I don't know which is right but I strongly suspect that whichever answer one arrives at with respect to the question of how many propositions the axiological impossibility results involve, this will be decided primarily by which of these approaches one takes. Are you a particularist? If so, then the answer will be 'a lot'. Ditto if you're an $\forall$ -phobe. If however you're a maximal generalist, the answer will be 'very few'. This is important. These general methodological questions will settle one's view *not only* with respect to how many propositions the axiological impossibility results involve, but with respect to all 'how many' questions – whatever the exact domain of enquiry - in moral philosophy or axiology. Particularists and $\forall$ -phobes will say 'lots', maximal generalists will say 'one or a few', moderates will say 'somewhere in between'. With this in mind, return to the challenge posed above. Our opponent asked:

Won't it be possible to find some single, general principle that covers all of (e.g.) (G) through (M) such that realists must reject only this one, single principle (and all entailments of it)? And by contrast, won't error theorists have to reject a great many general, plausible principles?

The answer is that the contrast presented in this quotation is unlikely. If one thinks it *is* possible to find some *single, general* principle that covers all of (G) through (M), then this one is almost certainly methodologically inclined towards generalism. And if one is this way inclined, then one will also presumably think that the axiological that error theorists reject can also be stated at a fairly high level of generality and so are fairly few in number. If on the other hand one thinks it is *not* possible to find some single, general principle that covers all of (G) through (M), then this one is almost certainly more of an $\forall$ -phobe. And if one is this way inclined, then one will also presumably think that the principles that error theorists reject cannot be stated at a fairly high level of generality and so are fairly large in number. In either case the number of propositions that the error theorist and realist respectively will have to accept counter-intuitive results with respect to may well be in lock-step with respect to their rough order of magnitude.

4.3 An Overall Case

We can now put the overall argument together. The basic claim is that in light of the axiological impossibility results realism has some consequences that are much more counter-intuitive than any of those of error theory. When coupled with a plausible principle for aggregating counter-

intuitive consequences of a theory, and with a realistic view of how many counter-intuitive consequences realists and error theorists respectively must stomach, the result is surprisingly strong for the error theorist. As stated above, the error theorist's big weakness is meant to be the counter-intuitiveness of the first-order consequences of her view, and the realist's big strength is meant to be her avoidance of any such counter-intuitiveness. I have argued that the gap is likely much narrower on this front than one might have expected. That's great news for error theorists.

Having put the point in this - I hope modest - way, I shall now see how far I can push it. Can we make an argument for thinking that error theory's first order consequences are actually less counter-intuitive than realism's? It would be bold, but let's try. Our first two premises are methodological. The first is a hedged general claim about the relative plausibility of theories of some common subject-matter:

- (1) For some subject matter, S, and two theories of S, T_1 and T_2 , if the consequences of T_1 with respect to S are less counter-intuitive than those of T_2 with respect to S, then - absent some debunking explanation of why this is - T_1 is, all else equal, a better theory than T_2 .

The second premise is a claim - based on 4.1 - about the aggregation of the counter-intuitiveness of the consequences of a theory:

- (2) The counter-intuitiveness of a theory's consequences is a result of aggregating both their number and their degree of counter-intuitiveness according to some balance where number does not straightforwardly outweigh degree.

We then appeal to two substantive claims. The first is a 'numerical lock-step thesis' from 4.2:

- (3) The number of counter-intuitive axiological consequences of error theory is most likely of the same rough order of magnitude as the number of counter-intuitive axiological consequences of realism.

These three premises together suggest that if we're adjudicating between error theory and realism, one will be more plausible than the other - all else equal - if the *degree* of counter-intuitiveness of its consequences is significantly lower. Here, as argued in section 3, things look

good for the error theorist. Unlike the realist she doesn't have to accept anything like The Anti-Pareto Principle. She only has to accept wide-scope negations and can reject completeness altogether. So:

- (4) The degree of counter-intuitiveness of at least some of the axiological consequences of realism is much greater than the degree of counter-intuitiveness of any of the equivalent consequences of error theory.

The result of (1)-(4) is a cautious case for:

- (5) Error theory is, all else equal, a better theory than realism.

This is a striking result: with respect to its first-order axiological consequences error theory is more plausible than realism. What about the 'all else equal' clause? Aside for the axiological consequences of the theory, we need to think about its metaphysical and epistemological commitments. Here it is *much* less controversial that error theory fares well in comparison to realism. Error theory is simple, explanatorily powerful, unified and makes none of the hairy metaphysics of realism. So if, as (4) states, error theory fares better than realism – *or even fares roughly equally* - with respect to its first-order consequences, then surely it is better overall; it wins on overall 'explanatory points'.¹³ So from (5) we get:

- (6) Error theory is a better theory than realism.

This is the bold conclusion promised above. Is it too bold? Probably. The inference from (1)-(4) to (5) is not watertight. Clearly, the move from (5) to (6) would also need more work (though that's really a different paper). Nevertheless, there is an interesting argument here – one that is not only provocative, but also plausible enough to be worth taking seriously. The weaker conclusion – that the gap between the counter-intuitiveness of the first-order consequences of realism and of error theory is much narrower than one might have expected – stands anyhow.

5. The Issue of Counter-Possibles

¹³ Here I use Enoch's (2011) terminology.

In this short section I consider a complication that I don't have an easy solution for. Consider propositions like:

(N) Undeserved pain is worse than pleasure.

Error theorists can't accept this. They think that:

(O) There are no properties of being better than.

And so, they must also think that:

(P) It is not the case that: Undeserved pain is worse than pleasure.

As noted earlier, error theorists should say that their commitment to claims like (P) is much less counter-intuitive than it might seem; it *doesn't* entail that undeserved pain and pleasure are equally good. Now there is an obvious move that I *could* have made earlier – but did not – to further bolster this error theoretic plea for innocence. I could have claimed that the error theorist should say:

(Q) If there were any axiological properties (which there aren't), then undeserved pain would be worse than pleasure.

On a first pass this is exactly the kind of thing that the error theorist should want to say. It would make the counter-intuitive first-order axiological consequences of her theory look *even less* troubling; it would make clear that the error theorist, unlike the realist, can respond to axiological impossibility results by, as it were, 'leaving everything as it is' as the first-order level and *merely* rejecting a false *meta*-axiological presupposition. Why, then, did I not make this move?

Following the orthodoxy, let's call a counterfactual with a necessarily false antecedent a 'counter-possible'. Now let's suppose – as the error theorist does – that error theory is true. Presumably if it is true, it is necessarily true. This means that if the error theory is true, then (Q) is a counter-possible. The truth value of counter-possibles is disputed. On one reading, they are vacuously true because their antecedents are necessarily false.¹⁴ If this reading is right, then while error

¹⁴ E.g. Lewis 1986, Williamson 2007.

theorists can truly make claims like (Q) they can't *really* claim that this renders their view any less counter-intuitive. This is because the very same reasoning that yields (Q) also yields the rather less rhetorically attractive:

(R) If there were any axiological properties (which there aren't), then pleasure would be worse than undeserved pain.

This shows that an error theorist who accepts the orthodox view of counter-possibles can't really use it to appeal to (Q); or rather, if she can, she can't claim that appealing to (Q) bolsters her credentials to have avoided implausible or counter-intuitive consequences.

There is, however, another way of thinking about counter-possibles. According to this alternative, counter-possibles can be non-vacuously true (or false). Their truth value is determined by the *impossible worlds* that they describe.¹⁵ To see how this works, consider the following example. Suppose that an error theory of number-ascriptions is true. Now consider the following propositions:

(S) If there were any numbers, then 2 would be greater than 1.

(T) If anything were any numbers, then 1 would be greater than 2.

According to the 'impossible worlds' theorist (S) is non-vacuously true and (T) is non-vacuously false. This is because at the nearest impossible world(s) in which there are any numbers, 2 is greater than 1. Let's just suppose for now that this is the right way of thinking about counter-possibles. We might think that an axiological error theorist could use it to great effect. She could claim that (Q) is non-vacuously true and (R) is non-vacuously false. This is exactly what she wants. It would make the degree of counter-intuitiveness of any claims to which she is committed look *even less* troubling than they already are. It would, as I put it above, make clear that the error theorist, unlike the realist, 'leaves everything as it is' as the first-order level and *merely* rejects a *meta*-axiological presupposition.

So it seems, then, that the impossible worlds approach to counter-possibles would be good for the error theorist. In fact, I don't think this is true. I think it would be very bad indeed for the error theorist; at least where the argument of this paper is concerned. Why? As we have already

¹⁵ E.g. Yagisawa 2010.

seen, the (e.g) Arrhenius axiological impossibility results show that the following cannot all be true:

- The Egalitarian Dominance Condition
- The Weak Non-Sadism Condition
- The General Non-Extreme Priority Condition
- The Weak Quality Addition Condition
- The Non-Elitism Condition

Let's suppose, as we have done so far, beginning in section 3, that The Egalitarian Dominance Condition is the most vulnerable of these. So both realist and error theorist must accept:

(A) The Egalitarian Dominance Condition is not true.

Now as we saw in section 3 above the realist must, as a consequence, also accept the highly counter-intuitive The Anti-Pareto Principle.

(F) 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 either worse than, equally as good as or on a par with B, other things being equal.

It is a strength of error theory, I argued, that it isn't also committed to this principle. But now suppose that the impossible worlds approach to counter-possibles is correct. Presumably, the error-theorist will now have to endorse a counter-possibly conditionalised version of every first-order claim that the realist will make. This will have to include The Anti-Pareto Principle! So the error theorist will be committed to:

(U) If there were any axiological properties, The Anti-Pareto Principle would be true.

This is *not* something that the error theorist should welcome. The point of appealing to counter-possible conditionals – like (Q) – was to make the counter-intuitive consequences of error theory look even less counter-intuitive; as though error theory merely commits us to denying a meta-axiological presupposition of otherwise plausible axiological claims. But if, in appealing to

counter-possibles, the error-theorist ends up accepting claims like (U), then she surely loses this advantage. She ends up accepting the same extremely hard to accept first-order claims as the error theorist (e.g. The Anti-Pareto Principle), just conditionalised on a necessary falsehood.

This leaves the error theorist biting her finger-nails. She should hope that the impossible worlds reading of counter-possibles is false and the orthodox reading is true. Otherwise, she loses some of her advantage; the advantage of being able to 'leave everything as it is' at the first-order level. I do not know whether the orthodox reading or the impossible worlds reading is the right one nor shall I hazard a guess here. We'll have to wait and see.

6. Conclusion

I began with a quotation from the closing sections of *Reasons and Persons*. It is also a good place to finish. Two of the defining features of Parfit's work are optimism about convergence in the moral judgment of practitioners, and metaethical realism. Axiological study since *Reasons and Persons* has given us cause to question the former. I have argued it therefore gives us cause to question the latter.

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 & 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 & E. Dzhafarov (Eds.), *Descriptive and normative approaches to human behavior: Advanced series on mathematical psychology*. Singapore: World Scientific Publishing Company.
- Arrhenius, Gustaf. 2012. Population Ethics and Metaethics. *Iride: Filosofia e Discussione Pubblica* 25 (1):35-44.
- Arrhenius, Gustaf. *Population ethics: The challenge of future generations*. Forthcoming with Oxford: Oxford University Press.
- Barcan Marcus, R. 1980. Moral dilemmas and consistency. *The Journal of Philosophy*, 77(3), 121–136.
- Blackburn, S. 1998. *Ruling Passions*. Oxford: Oxford University Press.
- Chang, R. 2002. The Possibility of Parity. *Ethics* 112(4): 659-688.

- Cowie, C. 2020. A New Argument for Moral Error Theory. *Nous*. Online first: <https://doi.org/10.1111/nous.12357>
- Enoch, D. 2011. *Taking Morality Seriously*. New York: Oxford University Press.
- Gibbard, A. 2003. *Thinking How to Live*. Cambridge, Mass.: Harvard University Press.
- Guttenplan, S. 1980. Moral realism and moral dilemmas. *Proceedings of the Aristotelian Society*, 80(1), 61–80.
- Lewis, D. 1986. *On the Plurality of Worlds*. Oxford: Blackwell.
- Mackie, J. L. 1974. Sidgwick's pessimism. *The Philosophical Quarterly*, 26(105), 317–327
- Ng, Y.-K. 1989. What Should We Do About Future Generations? Impossibility of Parfit's Theory X. *Economics and Philosophy* 5: 135–253.
- 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.
- Plantinga, A. 1974. *God, Freedom and Evil*. New York: Harper and Rowe.
- Shafer-Landau, R. 2003. *Moral Realism: A Defense*. New York: Oxford University Press.
- Schroeder, M. 2008. *Being For*. New York: Oxford University Press.
- Williams, B. 1965. Ethical consistency. *Proceedings of the Aristotelian Society*, 39(Supp.), 103–124.
- Williams, B. 1966. Consistency and realism. *Proceedings of the Aristotelian Society*, 11(Supp.), 1–22.
- Williamson, T. 2007. *The Philosophy of Philosophy*. Oxford: Oxford University Press.
- Yagisawa, T. 2010. *Worlds and Individuals, Possible and Otherwise*. Oxford: Oxford University Press.