(2016) *Philosophical Studies* 173(8): 2069-2086

Rossian Totalism About Intrinsic Value*

Luis R.G. Oliveira

Abstract: This paper defends a novel account of how to determine the intrinsic value of possible worlds. Section 1 argues that a highly intuitive and widely accepted account leads to undesirable consequences. Section 2 takes the first of two steps towards a novel account by clarifying and defending a view about value-contribution that is based on some of W. D. Ross' claims about the value of pleasure. Section 3 takes the second step by clarifying and defending a view about value-suppression that is based on Ross' claims about the interplay between different prima-facie duties. Section 4 states and defends the account that I call *Rossian Totalism*. According to this account, the atoms of intrinsic value within a world only sometimes contribute their intrinsic value to the value of that world.

Keywords: Metaethics, Intrinsic Value, Repugnant Conclusion, W.D. Ross

Introduction

Some of us would like to know what makes a possible world A better than a possible world B. Perhaps we would like to know this in order to better fulfill our moral obligations—we think that we ought to perform only those actions that are part of the best possible world available to us. Perhaps we believe that such knowledge could help us to better understand some of God's decisions—we think that God would only have created the best possible world. Perhaps we have even other reasons. In the sense of the *better-than* relation that is relevant to these concerns, however, knowing what makes a possible world A better than a possible world B requires understanding what is it that makes worlds *intrinsically valuable*—that is, what makes them valuable in and for themselves.[1] We thus need a method for calculating the intrinsic value of possible worlds. If we want to know what makes a possible world A better than a possible world B, we need to know (first) *what* contributes to the intrinsic value of possible worlds, and we need to know (second) *how* it does

*I am grateful to Christopher Meacham, Peter Graham, Ernesto Garcia, Kristian Olsen, Miles Tucker, Scott Hill, and especially Fred Feldman for discussion and comments on previous drafts.

[1] I am here following the standard assumption that a possible world A is *better than* (in the relevant sense) a possible B if and only if the total amount of intrinsic value in A is higher than the total amount of intrinsic value in B.

so. We need to know (first) the *sources* of intrinsic value, and (second) the *function* that takes us from these sources to the intrinsic value of a possible world. Only then will we be in a position to properly address philosophical concerns such as those outlined above.

To answer these 'what' and 'how' questions, in this case, is to produce an *axiological* theory—a theory of value. Here I make four guiding axiological assumptions. First, I take it that those who think that typical talk of intrinsic value is incoherent are mistaken.[2] Second, I take it that the bearers of intrinsic value are states-of-affairs, including maximally complex states-of-affairs such as possible worlds (cf. Chisholm 1968, 1975). Thus a kiss is not itself intrinsically valuable, but perhaps 'the state-of-affairs of being kissed by a loved one' is. Third, I take it that speaking coherently about the intrinsic value of various states-of-affairs requires the notion of *basic intrinsic value states* as the most fundamental bearers of intrinsic value (cf. Feldman 2000). A basic intrinsic value state is a pure attribution of a core intrinsically valuable property or relation that has a determinate and indefeasible intrinsic value. Thus 'the state-of-affairs of being kissed by a loved one', if intrinsically valuable, is either a basic intrinsic value state or somehow derives its intrinsic value from the basic intrinsic value states that occur within it (perhaps 'the state-of-affairs of feeling pleasure').[3] Fourth, I take it that we can numerically represent and perform mathematical calculations with the intrinsic values of intrinsically valuable states-of-affairs. Thus it makes perfect sense to say that 'the state-of-affairs of being kissed by a loved one' has 10 units of intrinsic value, and that 'the state-of-affairs of being hugged by a loved one' has 5 units of intrinsic value, and that therefore 'the state-of-affairs of being both kissed and hugged by a loved one' has 15 units of intrinsic

[2]See, for example, Peter Geach (1956) for a classic statement. The central worry (roughly) is that *goodness* (or perhaps the term 'good') is always relative to some category—good *knife* or good *cook*, for example—and never some all inclusive, non-relative, property (or term). Since typical talk of the intrinsic value of possible worlds is talk of some agent-, category-, purpose-neutral goodness (or is the deployment of the term 'good' in these ways), typical talk of intrinsic value, so the worry goes, is incoherent.

[3]The second and third assumptions seem individually plausible to me, and will play an important role in being precise about the problems and solutions discussed below. But they are not strictly required for my argument. What is required is some precise account of the bearers of intrinsic value in general and their relation to the atoms of intrinsic value in particular. I invite those who take the bearers of intrinsic value to be properties (cf. Butchvarov 1989), or facts (cf. Ross 1930, Lemos 1994), or states (cf. Zimmerman 2001), to make the relevant substitutions.

value.[4] Given these four guiding axiological assumptions,[5] we can reformulate the relevant 'what' and 'how' questions in the following way:

> **(a)** What kinds of basic intrinsic value states are there?[6]
>
> **(b)** How is it that basic intrinsic value states contribute to the intrinsic value of possible worlds?

An answer to (a) will tell us which kinds of states-of-affairs are the most fundamental sources of intrinsic value, and thus which kinds of states-of-affairs can contribute to the intrinsic value of possible worlds. An answer to (b) will tell us which function takes us from the values of the basic intrinsic value states that occur within a possible world to the intrinsic value of the possible world itself.

On one reading, however, part four of Derek Parfit's *Reasons and Persons* produces a severe constraint on what counts as an adequate answer to (a) and (b). According to this reading, Parfit has argued that any plausible answer to (a) and (b) has a particularly undesirable consequence: possible worlds that many of us think are obviously bad turn out to have what many of us think is a very high total amount of intrinsic value—an amount even higher than possible worlds that many of us think are obviously good (or at least obviously better). Parfit (1984, 388) refers to this result as "the repugnant conclusion":

> For any possible population of at least ten billion people, all with very high quality of life, there must be some much larger population whose existence, if other things are equal, would be better, even though its members have lives that are barely worth living.[7]

[4] Using Ruth Chang's (2002), terminology, this amounts to the assumption that intrinsic values (or at least those relevant below) are not *incommensurable*. This assumption, once again, is helpful but not strictly required for the problems to arise. What is required, instead, is that the relevant intrinsic values are not *incomparable*—to use Chang's terminology one more time. In making this fourth assumption, that is, I am in agreement with Chang in both taking incomparability as the most relevant notion of the two, and in not being moved by the arguments claiming that it is the case.

[5] These four assumptions are in line with what has been called the "Moorean concept of intrinsic value." See Bradley (2006).

[6] Or, more generally, how many different atomic core attributions of intrinsic values are there?

[7] Parfit (at least explicitly) seems to say nothing about possible worlds, intrinsic value, or answers to (a) and (b). But we can read him as making such claims. We can take Parfit to mean "possible world" by "possible population," and we can take his claims about the "better than" relation to be claims about the intrinsic values of the possible worlds under consideration. We can also take his talk of "quality of life" as designating a class of answers to (a) (to be discussed in detail below). This paper argues that if we take Parfit in this way, at any rate, we then take him as producing a hairy problem in axiology.

Intuitively, a world with ten billion people living great lives is better than a world with a much larger population living lives that are barely worth living. And it will be "repugnant," we can say with Parfit, if no plausible pair of answers to (a) and (b) can buttress this intuitive judgment. What I will call *Parfit's Constraint*, then, is the claim that successfully delivering this judgment is a minimal condition for the adequacy of any pair of answers to (a) and (b). If Parfit is correct, of course, then there are no such adequate answers. But the need for some such adequate answer has so far been unappreciated: we are in no position to make claims about what makes a possible world A better than a possible world B until we have an answer to (a) and (b) that respects Parfit's Constraint.

In this paper I argue that our best candidate for a minimally adequate answer to (a) and (b) takes a view I call Rossian Totalism as its answer to (b). In section 1, I briefly consider a highly intuitive answer to (b), which I call Feldman's Totalism. I argue that Feldman's Totalism produces the relevant undesirable consequences irrespective of which answer to (a) it is paired with. I also take a quick look at Moore's concept of an "organic unity," and I claim that though it does not have the resources to rescue Feldman's Totalism it nonetheless points the way towards an alternative answer to (b). In section 2, I take the first of two steps towards a novel answer to (b) by clarifying and defending a view about value-contribution (from a basic intrinsic value state to a world) that is based on some of W. D. Ross' claims about the value of pleasure. In section 3, I take the second step by clarifying and defending a view about value-suppression that is based on Ross' claims about the interplay between *prima-facie* duties. In section 4, I state and defend an answer to (b) that I call *Rossian Totalism*, and I argue that it has the resources to satisfy Parfit's Constraint.

1. Feldman's Totalism

Fred Feldman (1986, 218) defends the following widely accepted answer to (b):

> **Feldman's Totalism (FT):** The intrinsic value of a possible world is equal to the sum of the intrinsic values of the basic intrinsic value states that occur in that world.

As we will see, it is hard to find an answer to (a) that, when paired with (FT), produces a minimally adequate answer to (a) and (b).

Most discussions about intrinsic value focus on answers to (a). G. E. Moore (1903, 232-73), for example, suggested that the states-of-affairs of "loving our friends for their goodness," "loving beautiful objects for their beauty," "hating what is ugly or bad," and "compassion, or sympathy, for those who suffer pain" were all examples of intrinsically good states-of-affairs. W. D. Ross (1930, 140), for another example, considered four things as intrinsically good: "virtue, pleasure, the allocation of pleasure to the virtuous, and knowledge." William K. Frankena (1964, 63-77), for yet another example, suggested a long list composed of no less than 18 kinds of states-of-affairs plausibly taken to be intrinsically good, including such states-of-affairs as "having a true belief," "being free," and "being healthy." Strictly speaking, none of them distinguished between intrinsically good states-of-affairs that are basic intrinsic value states and intrinsically good states-of-affairs that derive their intrinsic value from the basic intrinsic value states that occur within them as intrinsic properties. But there are glimpses of this distinction in their writings.[8] More recently, Feldman (2004, 66) has suggested that "the intrinsic value of a life is entirely determined by the intrinsic values of the episodes of intrinsic attitudinal pleasure and pain contained in the life." He is explicit about taking such episodes as basic intrinsic value states.

We need not enter here in a discussion about which is the correct answer to (a). Let us instead represent the basic intrinsic value states within a world, whichever these may be, schematically as p_1, p_2, p_3..., p_n. Perhaps all p's are of the same kind, as in Feldman's view. Perhaps there are several different kinds of p's, as in Moore's, Ross', and Frankena's views. Either way, if (FT) is true, then the intrinsic value of a possible world is equal to the sum of the values of the p's that occur there. This suggestion, however, leads to undesirable consequences. Consider the following case:

[8] Ross (1930, 140) says, for example: "I am unable to discover anything that is intrinsically good, which is not either one of these or a combination of two or more of them." The suggestion here seems to be that there are other things (presumably states-of-affairs) that are intrinsically good as well, though not in a basic or fundamental way.

Parfit's Case

> Imagine that possible world A contains 10 million people whose lives are great, each containing 100 units of intrinsic value. Now imagine that possible world B contains 300 million people whose lives are barely worth living, each containing 5 units of intrinsic value.

Irrespective of which basic intrinsic value states we take to be contributing their values to the lives in A and B, the total sum of intrinsic value in B is higher than the total sum of intrinsic value in A. If (FT) is true, then B has a higher intrinsic value than A, and is thus better than A. This is just the kind of consequence that fuels Parfit's repugnance. Every life in B is barely worth living and every life in A is great. It seems obvious that A is better than B, and it will be repugnant if our axiology cannot suggest otherwise. No part of this argument, notice, depends on a particular answer to (a).[9]

Notice that (FT) is an *additive* account. A bit differently, Moore (1903, 79) claimed that "the value of a whole must not be assumed to be the same as the sum of the value of its parts." This amounts to saying not only that we should not assume (FT)—which is a claim about the intrinsic value of possible worlds—but also that we should not assume something like (FT) for any intrinsically valuable complex state-of-affairs, some of which fall short of being entire worlds.

Consider the state-of-affairs of "someone being conscious of a beautiful object". Moore claimed that, in isolation, the more fundamental states-of-affairs of "someone being conscious of something" and "a beautiful object existing" have a very low intrinsic value. Perhaps we can say that these are two basic intrinsic value states in Moore's axiology. But when these two states-of-affairs appear together in the form of "someone being conscious of a beautiful object," Moore claimed, they then form a complex state-of-affairs whose intrinsic value is not just the sum of their very low intrinsic values. The value of this particular "whole," to put it in Moore's words, is much higher than the sum of the values of its "parts". Lemos suggests another example in support of the existence of such wholes. Consider the state-of-affairs of "Jones being displeased that Smith is suffering". Lemos claims that, when

[9] In light of these worries, some have returned to Mill's distinction between *higher* and *lower* pleasures and suggested that perhaps certain basic intrinsic value states cannot contribute as much intrinsic value to a life as certain other basic intrinsic value states, no matter how many there are of the former kind in that life. See Mill (1863, 408-9), Ross (1930, 150), Lemos (1993, 482-4), and even Parfit (2004, 17-8), though he expresses doubts about the plausibility of this suggestion. Variations of Parfit's Case can be brought against this maneuver as well.

in isolation, the states-of-affairs of "Jones being displeased" and "Smith suffering" each has a negative intrinsic value. Perhaps we can say that these are two basic intrinsic value states in Lemos' axiology. But when these two states-of-affairs appear together in the form of "jones being displeased that Smith is suffering," Lemos (1998, 325-6) claims, they then form a complex state-of-affairs that is intrinsically good, a state-of-affairs that has a positive intrinsic value. If Lemos' judgments are correct, then the intrinsic value of this particular whole is much higher than the sum of the intrinsic values of its parts. Moore (1903, 87) suggests the name "organic unity" for such wholes whose intrinsic value is not equal to the sum of the intrinsic values of its parts. In the terminology of our present discussion, we can say that an organic unity is a complex state-of-affairs whose intrinsic value is not equal to the sum of the intrinsic values of the basic intrinsic value states that occur within it.

If organic unities exist, then perhaps there are even further answers to (a) that Moore, Ross, Frankena, and Feldman have not yet considered. But this is beside the point. It is easy to see that adding the values of the organic unities that occur within a world, together with the values of the basic intrinsic value states that occur within that world, produces the same undesirable consequences produced by adding just the latter. (One might even suggest that organic unities are themselves basic intrinsic value states, ones that have other basic intrinsic value states as their proper parts.) The existence of organic unities, that is, is no lifeboat for (FT).[10]

But perhaps possible worlds are *themselves* organic unities. That is, perhaps the intrinsic value of possible worlds is not equal to the sum of the intrinsic values of the basic intrinsic value states that occur within them. Indeed, it follows from a denial of (FT), and from the claim that possible worlds do have intrinsic values, that possible worlds *are*, in fact, organic unities. We have not gone so far as proving that (FT) is false, of course, but we have adduced good reasons for claiming that (FT) cannot be part of a minimally adequate answer to (a) and (b). Perhaps taking possible worlds as organic unities can provide us with such an answer. In fact, I think we can extract some such answer from some of W. D. Ross' claims about goodness and rightness. The rest of this paper is thus an attempt to draw exegetical and philosophical support for an answer to (b) that takes possible worlds as organic unities in a particularly Rossian way.[11]

[10]This is because (FT) precludes possible worlds from being organic unities themselves and, consequently, prevents the possibility of *defeat*. I will have more to say about defeat below.

[11]One alternative attitude at this point is simply to abandon Parfit's Constraint and embrace the repugnant conclusion. See Huemer (2008) for a defense of this line. I am here assuming that such

2. Rossian Value-Contribution

Ross (1930) notes an interesting worry with the notion that pleasure is intrinsically valuable: though pleasure seems to sometimes convince us that a certain state of affairs is good, at other times pleasure seems to fail to convince us in the same way. More exactly, while taking pleasure seems to be intrinsically valuable *in cases where the object of the pleasure is innocent*, pleasure seems not to be intrinsically valuable *in cases where the object of the pleasure is vicious*. This and other reasons led Ross to articulate a view of pleasure as only *prima facie* intrinsically good.[12] He says:

> Pleasure seems, indeed, to have a property analogous to that which we have previously recognized under the name of conditional or *prima facie* rightness. An act of promise-keeping has the property, not necessarily of being right but of being something that is right if the act has no other morally significant characteristic (such as that of causing much pain to another person). And similarly a state of pleasure has the property, not necessarily of being good, but of being something that is good *if the state has no other characteristic that prevents it from being good*. (Ross 1930, 138 my emphasis)

In an attempt to solve problems that arise independently of the repugnant conclusion, Ross here opens the door for the maneuver that, as I will argue, takes part in its solution.

There are, however, different ways of understanding what Ross himself means by his notion of *prima facie* goodness.[13] One way is to read Ross as suggesting that a complex state-of-affairs which contains as a part a basic intrinsic value state of 'someone taking pleasure in x' is not always itself intrinsically valuable. This reading suggests itself given Ross' analogy between *prima facie* goodness and *prima facie* rightness. For Ross, after all, what makes an action *overall* right (or what makes it our *all-things-considered* moral obligation, or our 'duty proper') is the *balance* of *prima-facie* rightness and wrongness. For any act-token A, there are features of A

a drastic measure should be reserved as a last resort. I take a similar line against the premature denial of the transitivity of the better-than relation (cf. Temkin (1987)).

[12]Other reasons include (i) the counter-intuitive fact that, given his belief that we have a prima facie moral obligation to bring about what has intrinsic value, it seems to follow that we have a prima facie moral obligation to bring about our own pleasure, and (ii) the fact that undeserved pleasures do not seem intrinsically valuable either. Notably, Ross (1939) later abandons his early view that pleasure is *prima facie* intrinsically good for the view that pleasure simply is not intrinsically valuable at all.

[13]See Stratton-Lake (2002, 122-30) for a discussion of Ross' views on this point.

that count as considerations of a certain strength for it being right, and there are features of A that count as considerations of a certain strength for it being wrong. What makes A overall right, then, is the balance of these considerations: if the sum of the strength of the rightness-considerations is higher than the sum of the wrongness-considerations, then A is overall right; otherwise it is overall wrong.[14] Perhaps, then, what Ross means by saying that pleasure is *prima facie* good is strictly analogous: pleasure is a goodness-consideration of a certain strength that is sometimes enough (or jointly enough) to make some complex state-of-affairs itself good, but sometimes not enough (nor jointly enough).[15] If we read him in this way, however, then we read him as saying nothing at all that is novel or insightful: on every view we have so far considered, complex states-of-affairs sometimes contain basic intrinsic value states with a positive intrinsic value without having a positive intrinsic value themselves—e.g., the state-of-affairs of a man taking incredible pleasure in something that, unbeknownst to him, is causing incredible suffering to other thousands. From the start we have assumed that positive basic intrinsic value states are *prima facie* good in this sense: they merely contribute to the goodness of a complex state-of-affairs of which they are a part of—and in this way are like the *prima-facie* duties that contribute to the rightness of an act-token—but are not themselves always capable (jointly or alone) of making that complex state of affairs itself good.[16]

But we can instead read Ross as suggesting that the basic intrinsic value state of 'someone taking pleasure in x' is not always *itself*, in some sense, valuable. There are two ways of making sense of this suggestion. The first way is to read Ross as abandoning a core Moorean thesis about intrinsic value, namely, that it supervenes on intrinsic properties. For given our understanding of basic intrinsic value states, two basic intrinsic value states of the same *kind*—say, someone taking pleasure in x—share the same intrinsic properties (cf. Feldman 2000).[17] Though on this read-

[14]Ross (1930) says very little about what affects the strengths of each of these considerations, other than that it varies with the circumstance. But Ross (1939) says a bit more, at least with respect to promise-keeping. See Olsen (2014) for a discussion.

[15]This is the reading that Stratton-Lake (2002) calls *atomistic*.

[16]Even worse, on this interpretation, Ross' notion of *prima-facie* goodness fails to solve the problem of vicious pleasures it was meant to solve: the problem is that pleasure in those cases seems to not contribute at all to the goodness of the complex state-of-affairs, instead of simply having its contribution outweighed.

[17]See Zimmerman (1980) for a defense of the claim that the object of pleasure is part of the intrinsic properties of basic intrinsic value states of pleasure. If he is right, then reading Ross in this way does not (at least not clearly) commit him to denying the supervenience of intrinsic value

ing Ross at least gathers the resources to solve the problem of vicious pleasures, he ends up saying something that simply makes hash of our guiding axiological assumptions—assumptions he himself took on board. The second way of understanding Ross' suggestion is to take him as making a very Moorean claim instead, namely, that while basic intrinsic value states (or at least states of pleasure) have their intrinsic value indefeasibly, the amount of intrinsic value that they *contribute* to a complex state-of-affairs of which they are part can vary.[18] So while basic intrinsic value states of pleasure are always intrinsically valuable, it is not always the case that they contribute that value to the complex state-of-affairs that have them as intrinsic properties. This is what it means for a basic intrinsic value state to be *prima-facie* good: to be capable of contributing its value, yet capable of failing to contribute it as well. In other words, this is the suggestion that a basic intrinsic value state's contribution of intrinsic value to a world can be *defeated* by other features of that world.

Leaving aside the important issue of exegesis, I will here suggest that we generalize this last interpretation of Ross' claims about the value of pleasure to every basic intrinsic value state. This will mean accepting the following thesis:

> **Rossian Value-Contribution (RVC):** Basic intrinsic value states are *prima-facie* intrinsic-value-contributors to the possible worlds in which they occur.

(RVC) is the first of two steps towards the answer to (b) that I will later call *Rossian Totalism*.

Recall that I have modeled my discussion around the axiological assumption that basic intrinsic value states are the fundamental bearers of intrinsic value. At the outset, I listed some of the characteristics of such entities, including the fact that their intrinsic value is *indefeasible*. This may seem in tension with my suggestion that a basic intrinsic value state can fail to contribute to the intrinsic value of a possible world, depending on other features that a possible world may or may not have. But there is no tension here. My suggestion is not that a basic intrinsic value state may or may not change *its* intrinsic value, depending on features of the possible world within which it occurs. Such suggestion would indeed be in tension with Feldman's claims. Instead, I am suggesting that whether or not a basic intrinsic value state *contributes* its intrinsic value to the possible world within which it occurs

on intrinsic properties.

[18] This is the reading that Stratton-Lake (2002) calls *holistic*. See Moore (1903) for the distinction between intrinsic value and contributory value.

depends on other features of that world. While a basic intrinsic value state has its intrinsic value of necessity,[19] it may or may not *contribute it* to the possible world within which it occurs. If a basic intrinsic value state does contribute its value, then I refer to it as a *de facto* intrinsic-value-contributor; if it does not, then I refer to it as a merely *prima-facie* intrinsic-value-contributor.

This suggestion, in fact, is not entirely new. Besides mirroring some of Ross' claims about the value of pleasure, it differs merely in degree from similar distinctions made in previous discussions about the intrinsic values of possible worlds (or, rather, discussions about the "goodness" of future "societies"). At least since Brentano, for example, it has been suggested that a basic intrinsic value state may contribute a certain amount of its intrinsic value to the *life* within which it occurs while contributing a different amount of its intrinsic value to the *world* within which it occurs. What brings about this asymmetry in contribution, just as in (RVC), is always a feature of the world and not a feature of the basic intrinsic value state.[20] My suggestion is in the same spirit. While the distinction I am suggesting between being a merely *prima-facie* intrinsic-value-contributor and being a *de facto* intrinsic-value-contributor differs from these other suggestions by saying nothing explicitly about the contribution of intrinsic value to someone's life, it is at the same time similar to them in claiming that a basic intrinsic value state's contribution to the intrinsic value of the possible world within which it occurs depends on other features of that world. As I call them, these are the *suppressing features* of possible worlds.

3. Rossian Value-Suppression

I take suppressing features to be axiologically relevant *extrinsic properties of states-of-affairs* that do not have intrinsic value. I can thus reaffirm the initial guiding assumption that intrinsically valuable states-of-affairs are either basic intrinsic value states or else derive their intrinsic value from the basic intrinsic value states that occur within them. If I am correct, of course, then it is possible for a state-of-affairs to derive its intrinsic value from only *some* of the basic intrinsic value states

[19] See Feldman (2000) and Bradley (2006) for a defense of this claim. It is not essential for my present purposes that this be the case; I assume it here merely to make explicit that (RVC) is not in tension with this traditional claim about intrinsic value.

[20] See Feldman (1997) for an exposition of Brentano's version of this distinction, and for Feldman's own version as well. See Hurka (1983) for a version often called the "diminishing marginal value" view, and see Kavka (1982) for a version often called the "critical level" view. I will have more to say about their views below.

that it contains; but it is never the case that it can derive its intrinsic value from something *other* than basic intrinsic value states. Suppressing features, then, do not affect a possible world by adding positive or negative intrinsic value; they simply, at times, prevent basic intrinsic value states from contributing their intrinsic values. Suppressing features, that is, are the causes of defeat.

I will here suggest four suppressing features of possible worlds. My first suggestion is similar to a suggestion made by Gregory Kavka (1982, 104) when talking about "the intrinsic moral desirability (or undesirability) of the existence of certain conditions of society or the world." As Kavka goes on to remark in the same passage, one such undesirable state of the world is "being a slaveholding society," and another is being "an overcrowded society (world)." The reason why these states of the world are undesirable, however, is that the lives within such states are barely worth living (or "restricted," as he puts it). This seems importantly right. We can build this "undesirability" intuition into our axiology, without sacrificing our guiding assumptions, if we accept as a suppressing feature the extrinsic property of:

> **($\mathbf{S}_1$)** Being a positive basic intrinsic value state in a life that is barely worth living.

Imagine a Kavkian "slaveholding" world where billions of slaves toil miserably under the hot sun. Imagine that, every now and then, amidst their pains and sufferings these slaves defy their circumstances and enjoy a good laugh. Assume, for the sake of argument, that this is an instance of the following basic intrinsic value state: (b_1) Individual I taking pleasure in p of intensity x and duration y at t. Let us accept that these intrinsic properties determine the intrinsic value that b_1 has of necessity. Now, it is a feature of the possible world in which b_1 occurs that b_1 is a positive basic intrinsic value state in a life that is barely worth living. As I see it, this extrinsic property of b_1, this feature of the possible world in which b_1 occurs, is axiologically relevant. It is a feature capable of suppressing b_1's contribution to the intrinsic value of the possible world in which it occurs. (I'll have more to say about *when* such suppressing takes place just below.)

My second suggestion is similar to a suggestion considered by Feldman (1997, 208). As he puts it, "the value for the world of an episode of bliss or misery is the value contained in that episode *adjusted for desert*" (emphasis original). As this suggestion has it, the amount of intrinsic value that a basic intrinsic value state contributes to a life may be different from the amount of intrinsic value that it

contributes to a world since the latter contribution is "adjusted for desert." This also seems importantly right. Consider, for example, a case where a vicious criminal lives a life of bliss. If we consider the welfare-relevant features of this case in isolation—namely, all the individual states of affairs of this vicious criminal taking pleasure or pain in p—we are perhaps forced to conclude (other things being equal) that the life of this vicious criminal is very good. But we are intuitively opposed to a positive assessment of a world filled with such people once we consider the fullness of the case, that is, once we consider that all of these pleasures and delights are undeserved. We can build this intuition into our axiology as well if we accept as a suppressing feature the extrinsic property of:

(S_2) Being an undeserved basic intrinsic value state.

For any contribution to an individual's welfare (or to that individual's lack thereof) we can ask whether that individual deserves that positive (or negative) intrinsic value. Just as before, while the intrinsic properties of these basic intrinsic value states determine the intrinsic value that these states have of necessity, it seems plausible that features of the world (extrinsic properties of the basic intrinsic value states) affect whether they contribute to the intrinsic value of the possible world within which they occur.

For my third suggestion, consider the case of a hard-working virtuous woman who lives in misery. Keeping other things equal, we can say that she deserves a very high balance of welfare. Suppose she receives a balance that is very low. Cases of "cosmic injustice" like this suggest yet another plausible suppressing feature of possible worlds:

(S_3) Being a positive basic intrinsic value state in a life that has less value than it deserves.

Notice that some, though not all, of the cases in which S_3 applies are also cases in which S_1 applies. There is then the possibility of suppressing over-determination, just as in Ross's view there often is rightness or wrongness over-determination (ϕ-ing, say, will save a baby's life, will fulfill a promise, and will fulfill a special obligation to progeny). If an individual deserved a balance of 100 units of intrinsic value and only received 90 (or 5, as we can imagine to be the case of our virtuous woman), then every positive basic intrinsic value state in that individual's life is a positive basic intrinsic value state in a life that has less value than it deserves. Each of these

basic intrinsic value states no doubt contribute (of necessity) to make her life better, but their extrinsic property of being in a life that has less than it deserves can affect whether they count towards making the world in which she lives a better world.

Lastly, consider the case of a banker secretly stealing part of her clients's money. Considering the welfare-relevant features of this case in isolation, once again, it is possible that there is a positive balance of intrinsic value: say the banker is enjoying an enormously high level of welfare; say the clients are enjoying a lower, but nonetheless enviable level of welfare. Cases like this reveal another plausible suppressing feature of possible worlds:

> **(S_4)** Being a positive basic intrinsic value state whose existence is causally connected to an undeserved negative basic intrinsic value state.

This explains how even if the banker is a virtuous person who is simply unaware that most of her money is stolen money (perhaps she has been mistakenly trusting a vicious accountant who, in order to protect his own stolen money, has turned his boss into an accomplice) we might still think that her welfare does not contribute to the value of the world.

These four suggestions are of course tentative and programmatic. They are merely intended to illustrate the nature of suppressing features: axiologically relevant extrinsic properties of basic intrinsic value states that (i) do not have positive or negative intrinsic value, and yet (ii) can affect whether the intrinsic value of a certain basic intrinsic value state makes a contribution to the intrinsic value of the world in which it occurs. Perhaps there are even other suppressing features.[21] Either way, my claim is that the intrinsic value of a possible world depends on the interplay of suppressing features such as these and whatever basic intrinsic value states exist in that world.

Notice that I've tried to register the indebtedness of my identification of these suppressing features to the views we find in Kavka, Feldman and Hurka. Is my view, then, merely the conjunction of their views, perhaps together with even others more. It is important to see that the answer here is 'no'. Previous suggestions have correctly identified features of possible worlds that can, as I say, suppress the contribution of intrinsic value from a basic intrinsic value state to the possible world within which it occurs—that is, features that can be the cause of defeat. In this

[21] Consider: (S_5) being a basic intrinsic value state that is not a product of a free action. The reply to the problem of evil that is often called "the free will defense" seems to me committed to some such claim.

way, my suggestion is indeed quite similar to them. Where previous suggestions have failed, however, is in claiming that these features suppress a contribution of value *whenever they are present.*[22] I now wish to deny this claim. Instead, I suggest that suppressing features may or may not suppress the contribution of intrinsic value from a basic intrinsic value state to a world, depending on whether this suppressing feature is *more axiologically relevant* or *more stringent* than that basic intrinsic value state. This means accepting the following thesis:

> **Rossian Value-Suppression (RVS):** Basic intrinsic value state b is a *de facto* intrinsic-value-contributor in w if and only if w contains no suppressing feature f that is more axiologically relevant than b.

(RVS) is the second of two steps towards Rossian Totalism.

Consider what Ross (1930, 41-2) says about the notion of stringency as it applies to his theory of right action:

> Every act therefore, viewed in some aspects, will be *prima facie* right, and viewed in others, *prima facie* wrong, and right acts can be distinguished from wrong acts only as being those which, of all those possible for the agent in the circumstances, have the greatest balance of *prima facie* rightness, in those respects in which they are *prima facie* right, over their *prima facie* wrongness, in those respects in which they are *prima facie* wrong... For the estimation of the comparative stringency of these *prima facie* obligations no general rules can, so far as I can see, be laid down... [The] sense of our particular duty in particular circumstances, preceded and informed by the fullest reflection we can bestow on the act in all its bearings, is highly fallible, but it is the only guide we have to our duty.

Once again, there are different ways to understand what Ross intended to convey. On one reading, Ross is here claiming that there are no principles governing the circumstantial stringency of our *prima facie* obligations. Here the idea would be that, sometimes, the fact that action A is an instance of promise-keeping weighs more heavily than the fact that a competing action B is an instance of benevolence; at other times, however, an otherwise identical situation could be such that the relevant weights are different. This seems to me not only an incorrect reading

[22] For objections to Hurka's and Kavka's view, see Arrhenius (2000). For objections to Feldman's view, see Arrhenius (2003). These objections are variations of what I called Parfit's Case in section 1.

of Ross, but, more importantly, an implausible suggestion in itself: it blatantly violates the supervenience of the normative on the non-normative. (RVS) better not be committed to something along these lines.

Equally implausible, at least as an interpretation of the passage above, is the suggestion that Ross is here claiming that we can assign fixed values to ways of being *prima facie* right and wrong. Here the idea would be that, for any set of alternative acts available at a time, we can determine the balance of *prima facie* rightness and wrongness of these acts by simply identifying the ways in which they are *prima facie* right and wrong—say, one for being an instance of promise keeping, one for being an instance of benevolence—and by assigning to these features their *fixed value*—perhaps promise keeping always counts less than benevolence. Such a reading ignores Ross' talk of "stringency" altogether: on certain situations, *prima facie* right-making feature x (say, being an instance of promise keeping) counts more than z (say, being an instance of benevolence); on other situations, however, z counts more than x. As David McNaughton (2002, 86) puts it, "where [duties] conflict we have to decide, *in each particular case*, which is here the weightiest... Ross's view commits him to the wholly plausible claim that the stringency of a duty can vary from one occasion to another" (my emphasis). So neither can we plausibly take Ross's theory of right action to be a complete rejection of governing principles, nor can we take it to be too rigid in its view of the interplay between right-making and wrong-making features. Ross's view, and (RVS), are to be found somewhere in between.

On a more plausible interpretation, then, Ross is claiming in the passage above that our duty in any given situation is to perform the available act with the greatest balance of *prima facie* rightness over *prima facie* wrongness, where such a balance is determined by principles governing the *circumstantial stringency* of these features in a way that is perhaps not fully accessible to us. It is not the case that there are no general rules governing the circumstantial stringency of a right-making feature, but rather that, for reasons of epistemic access, no rules can "be laid down" to guide us. This avoids the mistake made by the first implausible interpretation suggested just above. It is also not the case that each right-making or wrong-making feature always weighs the same, but rather that our "fullest reflection" must determine the relevant weights—their circumstantial stringency—in each particular case. This avoids the mistake made by the second interpretation.

(RVS), then, is the claim that a basic intrinsic value state with the extrinsic

property of, say, being 'being an undeserved basic intrinsic value state' will sometimes contribute its intrinsic value to the possible world in which it occurs and sometimes not. Just as sometimes act A being an instance of promise-keeping will be more stringent than act B being an instance of beneficence—promise-keeping, that is, will be more deontologically relevant in some cases—and sometimes not, sometimes the intrinsic value of a basic intrinsic value state will be more stringent than its suppressing extrinsic property—intrinsic value, that is, will be more axiologically relevant in some cases—and sometimes not. In both cases, we are hard pressed to lay down rules for this interplay. In both cases, we have nothing but our best reflection on particular cases as our guide to circumstantial relevance.

4. Rossian Totalism

In the last two sections I have tried to explain and motivate Rossian views of value-contribution and value-suppression: (RVC) and (RVS). We are now in a position to state the answer to (b) that has the resources to meet Parfit's Constraint.

> **Rossian Totalism (RT):** The intrinsic value of a possible world is equal to the sum of the intrinsic values of the *de facto* intrinsic-value-contributors that occur within that world.

Since (RT) does not claim that the intrinsic value of a possible world is *always* equal to the sum of the intrinsic values of the basic intrinsic value states that occur within that world, possible worlds (or rather *some* possible worlds) are organic unities if (RT) is true.[23] Given the exegetical considerations registered while discussing (RVC) and (RVS), it is not easy to say whether (RT) is something to which Ross, on reflection, would assent. The claim that it is *Rossian*, however, seems on point: on one hand, it builds on a plausible reading of some of Ross' claims about the value of pleasure; on the other, it mirrors in axiology the central aspect of the view that Ross held in normative ethics, namely, circumstantial stringency or relevance. If one is antecedently committed to Ross's theory of right action (or even if one simply feels a natural pull towards it), then (RT) should seem like a natural complement.

[23] One may find that cases of organic unities where the value of the whole is higher than the sum of the value of the parts are a bit mysterious. (See such examples from Moore and Lemos in the previous section.) Where, one may ask, is the value coming from? Whatever one thinks about this, however, possible worlds, according to RT, are not organic unities of this mysterious kind. The value of possible worlds, according to RT, is either identical to the sum of the value of its parts, or smaller than that value.

Here is then a full statement of (RT) and its implications. (RT) takes the axiologically relevant properties of possible worlds to be basic intrinsic value states and some of its extrinsic properties (the suppressing features). This is the analogue of Ross' view that an act can at the same time be *prima facie* right in some ways and *prima facie* wrong in others.[24] I have suggested four plausible suppressing features just above. If, as Ross puts it, when "informed by the fullest reflection" we conclude that there are suppressing features in a possible world that are more stringent or, as I have put it at times, more axiologically relevant than certain basic intrinsic value states that occur within that world, then those basic intrinsic value states fail to qualify as *de facto* intrinsic value contributors. This is intended as the exact same operation by which we determine that, in a certain circumstance, keeping a promise is more stringent than doing some bit of good. The suppressed basic intrinsic value states are still there, of course, and continue to have the intrinsic value had by necessity in virtue of its intrinsic properties, and continue to contribute to the value of the life of which they are a part of—much like, on Ross's theory of right action, an outweighed *prima facie* duty does not cease to be a *prima facie* duty and often continues to require attention from the agent (for reparations, for example).[25] But the world within which these suppressed basic intrinsic value states occur is no better because of them. Yet none of this is to say that there are no rules governing the interplay of basic intrinsic value states and suppressing features. Rather, it is simply to say that, for reasons of epistemic access, no general rules can be laid down, and that we must do our best to determine relevance by reasoning and reflecting on a case by case basis.

In fact, it seems to me that such reasoning and reflection reveal that (RT) avoids the undesirable consequences faced by (FT) in section 1. Notice that (RT) retains the *spirit* of (FT): when we calculate the intrinsic value of a possible world, we still sum the intrinsic values of basic intrinsic value states. The difference, of course, is that (RT) clarifies the need to determine, before any calculation, which basic intrinsic value states are to be included in the sum—that is, which of the *prima*

[24]The analogy, of course, is not perfect. According to Ross, what the determines the overall rightness of an act is the interplay of, in a sense, the same kind of thing: ways in which the act is *prima facie* right or wrong. The related view in axiology perhaps should be that what determines the overall goodness of a state of affairs is the interplay of, in a sense, the same kind of thing as well: basic intrinsic value states. This disanalogy, I think, is quite alright. Here (RT) retains the spirit but not the letter of Ross' view.

[25](RT) does not imply, therefore, that suppressed basic intrinsic value states have no normative significance.

facie intrinsic value contributors are also *de facto* intrinsic value contributors. It is at the point of this determination that suppressing features play their roles. If they are found, on reflection, to be more stringent (or more axiologically relevant) than certain basic intrinsic value states, then they suppress the contribution of intrinsic value from these basic intrinsic value states to the world in which they occur. For an evaluation of (RT) against objections, then, it must suffice if we can record our considered judgments of relevant cases as cases where the suppressing features seem to be, on the strength of our fullest reflection, more stringent than certain basic intrinsic value states. This should be no surprise: evaluation of Ross' theory of right action also turns on recording our judgment of certain cases as cases where we judge, on reflection, that our *prima-facie* duty of "beneficence" is more stringent than our *prima-facie* duty of "fidelity," for example, even without identifying an underlying rule that delivers this result in a systematic way.

Consider Parfit's Case again. It says that possible world A contains 10 million people whose lives are great, each containing 100 units of intrinsic value. If we wish to evaluate the intrinsic value of A according to (RT), however, then the details of A are so far under-specified. It is not enough to simply list the basic intrinsic value states that occur within A. According to (RT), basic intrinsic value states are only *prima facie* intrinsic value contributors to the possible worlds in which they occur. In order to calculate the intrinsic value of A, we must first specify parfit's Case in a level of detail that allows us to determine whether there are any suppressing features in A that are more stringent than some (or perhaps all) of the basic intrinsic value states that occur in A. For our purposes, let us assume that A does not have any such feature. Let us assume that A is a truly great world, a world where no one is living a life that is barely worth living, a world where no one is undeserving of the positive basic intrinsic value states that occur in their lives, and so forth. A is thus a world where every basic intrinsic value state contributes to the intrinsic value of A. Since each life in A contains 100 units of intrinsic value, (RT) tells us that the intrinsic value of A is 100 x 10 million, or the sum of the intrinsic values of all the basic intrinsic value states that occur in A.

But Parfit's Case also says that possible world B has 300 million people whose lives are barely worth living, each containing 5 units of intrinsic value. If we here assume that every basic intrinsic value state is also a *de facto* intrinsic value contributor, then we are forced to conclude, as before, that B has a higher intrinsic value than A. But according to (RT), before we can make any such judgment we

must once again specify Parfit's Case in a level of detail that allows us to determine whether there are any suppressing features in B that are more stringent than some (or perhaps all) of the basic intrinsic value states that occur in B.[26] In this case, we can immediately see that it is a feature of B that every positive basic intrinsic value state that occurs in it has the extrinsic property of being a positive basic intrinsic value state that occurs in a life that is barely worth living. This is the suppressing feature I called (S_1). More importantly, for every basic intrinsic value state in B, its extrinsic property of being a positive basic intrinsic value state that occurs in a life that is barely worth living, it seems to me, is more axiologically relevant for that world than the basic intrinsic value state itself. (I here invite the reader to consult his or her own intuitions.) As (RT) will have it, then, B is a world filled with *prima facie* intrinsic value contributors, but lacking even a single *de facto* intrinsic value contributor. Though B is filled with positive basic intrinsic value states, and filled with lives that have some value for those living it, B as a world is not worth anything at all. According to (RT), then, A is better than B.[27]

Once again, these suggestions are programmatic. What (RT) provides is not a clear-shut solution to the repugnant conclusion but rather an insight into how it might be solved. Compare this to the way in which Ross' theory of right action can help explain why we do not have an overall duty to help five strangers instead of helping a friend to whom we promised help; the Rossian here can only appeal to an intuition regarding an evaluation of stringency, and can only suggest that the normative facts, on their view, can plausibly be this way. My suggestions here come in the same spirit.[28]

[26]Note that this is not a violation of Parfit's "other things being equal" as long as the suppressing features are features that simply cannot be kept equal.

[27]Notice that the same reasoning also avoids what has been called "the mere addition paradox" (cf. Parfit 1984, 419-441). At some point along the continuum of worlds A, B, C,..., Z, the basic intrinsic value states that occur within the worlds get suppressed. In the same way, this reasoning resists Huemer's (2008) concessive defense of repugnance, and avoids the need for Temkin's (1987) denial of transitivity.

[28]One may worry that (RT) cannot explain the difference in value between two pretty bad worlds where one is nonetheless intuitively less bad than the other. There are two maneuvers available to (RT) in this case. One is bitting the bullet and claiming that (RT) correctly captures one sense in which both of these worlds are worthless. No one should want to bring these worlds about; no world should be brought about at all if these are the only options; and so forth. I am sympathetic to this maneuver. Another is supplementing (RT) with a notion of partial suppression. I am less sympathetic, though not opposed, to this maneuver.

References

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

Arrhenius, G. (2003). "Feldman's Desert Adjusted Utilitarianism and Population Ethics," *Utilitas: A Journal of Utilitarian Studies* 15(2): 225-236.

Bradley, B. (2006). "Two Concepts of Intrinsic Value," *Ethical Theory and Moral Practice* 9 (2): 111-130.

Butchvarov, P. (1989). *Skepticism in Ethics*. Indianapolis: Indiana University Press.

Chang, R. (2002). *Making Comparisons Count*. Routledge.

Chisholm, R. (1968). "The Defeat of Good and Evil," *Proceedings and Addresses of the American Philosophical Association* 42: 21-38.

Chisholm, R. (1975). "The Intrinsic Value in Disjunctive States of Affairs," Noûs, 9: 295-308.

Frankena, W. K. (1964). *Ethics*. Englewood Cliffs, NJ: Prentice Hall Inc.

Feldman, F. (1986). *Doing the Best We Can: An Essay in Informal Deontic Logic*. Dordrecht, Netherlands: D. Reidel Publishing Company.

Feldman, F. (1997). "Justice, Desert, and the Repugnant Conclusion," in Feldman, F. (ed) *Utilitarianism, Hedonism, and Desert: Essays in Moral Philosophy*, 195 -214. New York: Cambridge University Press.

Feldman, F. (2000). "Basic Intrinsic Value," *Philosophical Studies* 99 (3): 319-346.

Feldman, F. (2004). *Pleasure and the Good Life: Concerning the Nature, Varieties and Plausibility of Hedonism*. Clarendon Press.

Geach, P. (1956). "Good and Evil," *Analysis* 17: 3342.

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

Hurka, T. (1983). "Value and Population Size," *Ethics* 93 (3): 496-507.

Kavka, G. (1982). "The Paradox of Future Individuals," *Philosophy and Public Affairs* 11 (2): 93-112.

Lemos, N. (1993). "Higher Goods and the Myth of Tithonus," *Journal of Philosophy* 60 (9): 482-496.

Lemos, N. (1994). *Intrinsic Value*. Cambridge: Cambridge University Press.

Lemos, N. (1998). "Organic Unities," *Journal of Ethics* 2 (4): 321-337.

McNaughton, D. (2002). "An Unconnected Heap of Duties?". In Phillip Stratton-Lake (ed.), *Ethical Intuitionism: Re-Evaluations*: 76-92. Oxford University Press.

Mill, J. S. (1863/1962). *Utilitarianism*. Cleveland, World Pub. Co.

Moore, G. E. (1903/2004). *Pricipia Ethica*. Dover Publications.

Olsen, K. (2014). "Ross and the Particularism/Generalism Divide". *Canadian Journal of Philosophy* 44 (1): 56-75.

Parfit, D. (1984). *Reasons and Persons*. Oxford: Oxford University Press.

Parfit, D. (2004). "Overpopulation and the Quality of Life," In J. Ryberg & T. Tannsjo (eds.), *The Repugnant Conclusion*, 7-22. Kluwer Academic Publishers.

Ross, W. D. (1930/2002). The Right and The Good. Oxford: Clarendon Press.

Ross, W. D. (1939/2000). Foundations of Ethics. Oxford: Oxford University Press.

Stratton-Lake, P. (2002). "Pleasure and Reflection in Ross's Intuitionism". In Phillip Stratton-Lake (ed.), *Ethical Intuitionism: Re-Evaluations*: 113-36. Oxford University Press.

Temkin, L. (1987). "Intransitivity and the Mere Addition Paradox," *Philosophy and Public Affairs* 16 (2): 138-187.

Zimmerman, M. J. (1980). "On the Intrinsic Value of States of Pleasure," *Philosophy and Phenomenological Research* 41 (1-2): 26-45.

Zimmerman, M. J. (2001). *The Nature of Intrinsic Value*. Lanham: Rowman and Littlefield.