

Antiprioritarianism

Hilary Greaves

University of Oxford

Abstract

Prioritarianism is supposed to be a theory of the overall good that captures the common intuition of “priority to the worse off”. But it is difficult to give precise content to the prioritarian claim. Over the past few decades, prioritarians have increasingly responded to this ‘content problem’ by formulating prioritarianism not in terms of an alleged primitive notion of quantity of well-being, but instead in terms of von Neumann-Morgenstern utility. The resulting two forms of prioritarianism (which I call, respectively, “Primitivist” and “Technical” prioritarianism) are not mere variants on a theme, but are entirely distinct theories, amenable to different motivating arguments and open to different objections. This paper argues that the basic intuition of “priority to the worse off” provides no support for Technical Prioritarianism: *qua* attempt to capture that intuition, the turn to von Neumann-Morgenstern utility is a retrograde step.

1 Introduction

Textbook utilitarianism comprises three components: a particular account of individual well-being (hedonism), a particular account of the relationship between individual well-being levels and the overall goodness of the state of affairs (the additive method of aggregation), and a particular account of the relationship between goodness of states of affairs and what one ought to do (maximising consequentialism). Each of the three components is highly controversial. I shall set aside the issues of hedonism and maximising consequentialism: this paper is concerned entirely with the utilitarian’s aggregative claim. (For brevity, I will usually use the shorter term ‘utilitarianism’, but I will always mean only the aggregative component.)

The distinctive feature of utilitarian aggregation is its notorious indifference to distribution: according to the utilitarian theory of the good, two states of affairs that agree on the sum total of utility are equally good, regardless of how the utility is distributed among people. But surely, one might think, the distribution is important too? Quite aside from any issues about hedonism or consequentialism, this basic thought is responsible for driving many away even from utilitarianism’s aggregative claim (for example, (Rawls, 1972, esp. p. 26), (Sen, 1973, p. 16), (B. Williams, 1973, pp. 142–3), (Scheffler, 1982)).

Prioritarianism is a view of aggregation that is supposed to accommodate this widespread intuition regarding the importance of distribution. While there is no canonical statement of the view, the headline slogan of ‘priority to the worse off’ captures its general spirit. Prioritarians would sign up to Nagel’s (1979) insistence that faced even with a choice between delivering a given benefit to one child and delivering a “substantially greater” benefit to a second child who is already better off, it might be “more urgent to benefit the [worse off] child”. In his classic article ‘Equality and Priority’ (1997), Parfit states the view as being that “benefitting people matters more the worse off they are”; in his recent (2012), it is “we have stronger reasons to benefit people the worse off they are”.

Taken on their own, there is clearly some truth, and an important truth, in these statements. The statements as they stand, however, are of course rather vague, and not clearly equivalent. Some work needs to be done to give expression to the prioritarian intuition in precise theory. One crucial step is in making precise the notion of ‘same size benefit’: under what conditions does one benefit accruing one person count as being the ‘same size’ as a different benefit accruing to a different person?

In answering this question, there has been an unannounced shift. In the early days, prioritarians held out hope for a sufficiently precise notion of benefit *size* that was not (however) reducible to any betterness *ordering* (either betterness-simpliter, or betterness-for-the-individual). Critics, however, questioned whether there really was any notion of benefit size that is irreducible to ordinal matters but nevertheless determinate enough to give content to the prioritarians’ claim. In response to this criticism, in the more recent literature, prioritarians have tended to stipulate instead that ‘same size benefit’ means *same size increase in von Neumann-Morgenstern utility*, where the latter notion is made precise using the technical resources of decision theory.

As a result of this shift, the subject has been changed. Early prioritarian writings were defending what I shall call ‘Primitivist Prioritarianism’; the recent literature concerns instead ‘Technical Prioritarianism’. I will argue that those whose driving motivation is the basic prioritarian intuition should not have embraced this shift: Technical Prioritarianism receives no support from that intuition. One way of arguing for this claim would be simply to point out the implausibility of insisting that such a general intuition is really about such a theoretical notion as that of von Neumann-Morgenstern utility. I endorse this line of argument; nevertheless, the present paper is addressed to those modern prioritarians who are not yet convinced by it. Here, I show that the method that modern prioritarians use to turn their intuition into a theory, when applied to a second intuition (concerning caution in the face of risk) that is just as plausible as and of a similar character to the basic prioritarian intuition, leads to a theory (“Technical Cautionism”) that is diametrically opposed to Technical Prioritarianism. Since neither intuition is itself entirely without merit, it must (even by the Technical Prioritarian’s own lights) be that the Technical Cautionist’s theory-building method is faulty; but then the Technical Prioritarian’s method must be faulty too, since it is the same method.

The structure of the paper is as follows. Section 2 describes the path from the prioritarian’s basic intuition to a core of the Technical Prioritarian theory. As above, I will *note* the implausibility of thinking that Technical (as opposed to Primitivist) Prioritarianism is supported by direct intuitions of the sort that the prioritarian appeals to, but I will assume for the sake of argument that the modern prioritarian is unconvinced. I will also *note* some uncomfortable choices, related to Harsanyi’s celebrated (1955) aggregation theorem, that the Technical Prioritarian has to make in effecting any extension of her theory to cover scenarios involving risk, but again I will suppose that my interlocutor is not convinced that these are problematic. Section 3 develops “antiprioritarian theory”: I state the “cautionist” intuition that I take to be analogous to the prioritarian intuition, and develop it into a theory in a manner that parallels section 2’s treatment of the prioritarian intuition. This will involve distinguishing between Primitivist Cautionism and Technical Cautionism; since I seek the analogue of Technical Prioritarianism, I will focus thenceforth on the latter. Section 4 notes that Technical Prioritarianism and Technical Cautionism contradict one another.

All this is by way of setup; the central *argument* of the present paper is in section 5, in which I argue, from the incompatibility between Technical Prioritarianism and Technical Cautionism, to the conclusion that the basic prioritarian intuition provides no support for Technical Prioritarianism. Section 6 argues that, while the paper’s direct concern has been with *prioritarian* attempts to capture the importance of distribution, similar issues apply to egalitarianism. Section 7 summarises.

The paper’s three appendices investigate the extension of the basic Technical Prioritarian and Technical Cautionist claims to cases involving risk in more detail. Appendix A discusses three possible extensions of Technical Prioritarianism, and surveys the ways in which each violates the assumptions of Harsanyi’s aggregation theorem. Appendix B performs the analogous tasks for Technical Cautionism. Appendix C collects together the various value functions that express the utilitarian position, the three versions of Technical Prioritarianism and the three versions of Technical Cautionism discussed herein, for ease of reference.

2 Technical Prioritarianism

2.1 Clarifications

The prioritarian seeks, then, to give expression to her intuition of ‘priority to the worse off’ in a precise theory. Three preliminary clarifications are in order.

First clarification: Let *Resource Prioritarianism* be the thesis that faced with a choice between delivering a given amount of concrete *resource* — say, ten sacks of rice, or a thousand dollars — to a badly off person or to a well-off person, it is, *ceteris paribus*, better (or it is more urgent, or it matters more, or we have stronger reason) to give the resource in question to a destitute

farmer than to someone who is already a millionaire. Resource Prioritarianism is uncontroversial. The choice just described, however, is a choice between giving a relatively small benefit to the millionaire (what's *he* going to do with yet another thousand dollars?) and giving a much larger benefit to the farmer. In particular, the unanimous Resource-Prioritarian judgment in this case is one the utilitarian is perfectly well able to agree with: in his terms, the point is that resources have diminishing marginal utility. To be at all distinctive, the prioritarian thesis must therefore be that it is better (or more urgent, etc.) to benefit the worse off, even if the size of benefit delivered is independent of whether one chooses to benefit the better-off or the worse-off person.

Second clarification: prioritarians do not advocate *absolute* priority to the worse off. If one is choosing between delivering a truly tiny benefit to a badly off person and delivering a much bigger benefit to a better off person — say, giving one extra grain of rice to the farmer or curing the millionaire of cancer — it may yet be better (more urgent, etc.) to do what one can for the better off person than to do the very little one can for the worse off person. (That is, prioritarians are not advocates of a maximin or leximin approach.) The ideology of 'size of benefit' is therefore crucial: it is only in cases in which size of benefit is equal that prioritarians will necessarily claim that it is better to benefit the worse off.

Third clarification: my focus throughout is on *axiological* prioritarianism. That is, I take the issue to be the betterness ordering of states of affairs, and the prioritarian's intuition to be that relative to any status quo, *the state of affairs that would result* from delivering a benefit of a given size to a worse off person is *better* than the one that results from delivering a benefit of the same size to a better off person. This axiological prioritarianism is to be distinguished from a *merely deontological* prioritarianism, which claims that if one is in such a choice situation then one ought (one has stronger reason) to deliver the benefit to the worse off party, but refrains from making any claim about whether or not this is because the resulting state of affairs is better. It is also to be distinguished from a *merely emotional* prioritarianism, which refrains from making any claims about which states of affairs are better *or* any claims about what one ought to do, and merely asserts something about what it is appropriate to *feel* (a greater sense of urgency, perhaps) when contemplating the delivery of certain benefits.

This choice of focus, of course, does not presuppose a consequentialist account of the relation between betterness facts and ought-facts. The point is familiar: any remotely sane normative theory, consequentialist or otherwise, must take the identification of the correct theory of the good to be at least *an important part* of the full moral story, whether or not the rest of the story supervenes on that part. Progress will be best served by first settling the betterness ordering, and reserving discussion of what one ought to do in the light of this

ordering for a later stage.^{1,2}

2.2 The prioritarian's 'same size benefit' relation

What, then, is the betterness ranking of states of affairs, according to axiological prioritarrians?

To express the partial answers to this question with which this paper is concerned, we need to consider *multiple* betterness orderings. For each individual i , there is a betterness ordering \succ_i : $A \succ_i B$ iff state of affairs A is better *for person i* than state of affairs B is. (Hedonists, for instance, hold that A is better than B for person i iff i enjoys a greater balance of pleasure over pain in A than in B.) In addition to this multiplicity of personal betterness relations, we also have an impersonal betterness ordering \succ : $A \succ B$ iff A is better than B *simpliciter* ('from the point of view of the universe').

Any such ordering (subject to technical conditions that will not concern us here; see (Debreu, 1960)) can be represented by a value function: an assignment of real numbers to states of affairs, such that one state of affairs is assigned a higher number than another iff the first is better than the second. Thus we have an overall value function, V , representing \succ ; and for each person i , we have an individual value function V_i , representing \succ_i .

The locus of disagreement between prioritarian and utilitarian is the relation between the overall value function V and the collection of individual value functions $\{V_i\}$. *As a first pass*, the disagreement is that whereas the utilitarian holds that overall value is just the sum of individual values —

$$V^u = \sum_i V_i \tag{1}$$

— the prioritarian insists that overall value is given instead by

$$V^p = \sum_i f(V_i), \tag{2}$$

¹It is not always clear whether prioritarrians have in mind axiological, deontological or fitting-feeling claims, but I do not think the axiological prioritarian is a straw man. Parfit (1995, section VII) clearly draws the distinction (in his terminology, between 'telic' and 'deontic' prioritarianism), stating that 'for most of my discussion, this difference does not matter'; his (2012) defends 'the telic priority view', and his statement of that view explicitly includes the clause that 'it would in one way *make the outcome better*' (my emphasis) if one acted so as to give people 'a greater sum of weighted benefits'. Williams (2012) explicitly argues for deontological as opposed to axiological ('teleological') prioritarianism, in response to objections that he takes to affect only the latter.

²An anonymous referee has suggested that the notion of better-simpliciter is unacceptably mysterious until and unless it have been given an analysis, perhaps in deontic terms (for example, 'more choiceworthy'). I don't agree, but those who do are free to substitute their preferred analysis throughout (any such analysis had better not collapse the axiological/merely-deontological distinction altogether). I leave it to the reader to examine whether or not any of the details of the arguments in this paper are affected by the particular substitution they prefer.

where f is a strictly increasing and strictly concave function. This transform f expresses the prioritarian's desired 'priority to the worse off': a given increase in V_i amounts to a greater increase in $f(V_i)$ if delivered to a person whose existing V_i is lower.

A moment's thought, however, reveals that we have not yet pinned down a genuine disagreement between prioritarian and utilitarian. The reason is that we have not yet placed enough restrictions on the individual value functions V_i : we have said only that they correctly represent the respective individual betterness orderings \succ_i . This leaves open a large class of functions V_i for a given ordering \succ_i , related to one another by arbitrary monotonically increasing transformations. For all we have said so far, therefore, it is open to the utilitarian to respond that his V_i just is the prioritarian's $f(V_i)$ — that is, that he encodes into the individual value functions all the 'priority to the worse off' that the prioritarian desires.

In search of a genuine disagreement, we might try further to restrict the V_i by insisting that a given increase in value of V_i must always correspond to the *same size of benefit* for the individual concerned. But this, of course, just shifts the question, and forces us to face the issue of what exactly is meant by 'same size benefit'.

One possibility is simply to *define* size of benefit as: size of contribution that this improvement to this person's lot makes to the overall goodness of the state of affairs. This definition, however, makes utilitarianism trivially true, prioritarianism trivially false; it is therefore unavailable to the prioritarian.

In the early days of prioritarian theory, prioritarians supposed that there was simply a *primitive* fact of the matter as to whether one benefit to one person was the 'same size' as a second benefit to a second person — a fact that is in itself entirely independent of any betterness ordering — and that these facts could be accessed by some sort of intuition. Those of a more positivist inclination complained that if the alleged same-size-benefit facts could not be read off from observed dispositions to choice behaviour, there could be no such facts. Defenders of prioritarianism retorted that since positivism is an excessively restrictive philosophy in general, this complaint has no force. (Both the primitivist position, and this sort of response to the positivist complaint, are explicit in (Sen, 1976, pp. 249–50).)

This dismissal of the positivist critique, however, was too quick. Granted that the positivists went too far, there clearly is something to the insistence that a postulated piece of ideology must have some connections to some other notions if it is to have any content.³ On the epistemic side: granted the validity of introspection as a method in general, if one is asked to consult one's introspection on the question of whether one benefit counts as being the 'same size' as another but is told nothing at all about the criteria of application of that term, one should complain that one does not know what it is one is supposed to be introspecting. The prioritarian's notion of benefit size, independent of size

³This weakening of the positivist position is most explicit in conceptual role semantics, but even those who reject that approach to semantics must acknowledge the point in some form.

of contribution to overall good, is in danger of being contentless. This criticism is pressed, in particular, by John Broome (e.g. (Broome, 1991, pp. 146–8 and sec.10.3)); I shall refer to it as ‘the content problem’ for prioritarianism.

In the more recent prioritarian literature, prioritarians have responded to the content problem by borrowing the resources of decision theory, and appealing to the notion of von Neumann-Morgenstern utility. The idea is well-known, but will be worth rehearsing nonetheless. The key observation is that while the task of representing an individual betterness ordering of *states of affairs* does not pick out one positive affine family of value functions as privileged over any other, matters are different once one incorporates the treatment of risk — once one turns, that is, to the task of representing an individual betterness ordering of *prospects*, i.e. probability distributions over states of affairs. The representation theorems of decision theory show that for any (individual betterness) ordering of prospects satisfying certain reasonable axioms, that ordering can be represented by an assignment of numbers (‘utilities’) to states of affairs, such that one prospect is ranked as better-for-the-individual than another iff the first has higher expected utility than the second, expected utility being a probability-weighted sum of utilities. In honour of two of the founding fathers of decision theory (von Neumann & Morgenstern, 1944), we call these numbers von Neumann-Morgenstern (vNM) utilities. Crucially, the representation theorem guarantees that vNM utilities are unique up to positive affine transformation. Given this feature, if the utilitarian advocates the overall value function

$$V^u(A) = \sum_i u_i^{VNM}(A) \quad (3)$$

while the prioritarian advocates

$$V^p(A) = \sum_i f(u_i^{VNM}(A)), \quad (4)$$

the two parties have a genuine disagreement. The value function (4) is the core claim of the position I shall call Technical Prioritarianism; in terms of our preceding discussion, it amounts to taking ‘same size benefit’ to mean ‘same size increase in von Neumann-Morgenstern utility’.⁴

The term ‘Technical Prioritarianism’ is my own, but it is chosen carefully. Although its advocates do not emphasise (and, I think, generally do not notice) this, *it is a technical claim*, in the sense that its central notion of von Neumann-Morgenstern utility is itself a technical one: it is not a primitive notion of quantity of well-being, but is implicitly defined by the theory of *ex ante*

⁴As is well recognised, if our starting point is an *independent* better-for-*i* ranking for each person *i*, this representation theorem does not yet give us any interpersonal unit comparisons, i.e. comparisons between the difference between A and B for person *i*₁ and the difference between C and D for a *distinct* person *i*₂: it merely gives us a separate unit-comparable utility scale for each person. This, of course, is why the issue of ‘interpersonal utility comparisons’ in particular has remained a vexed one after the advent of expected utility theory. Without undertaking a full survey of the possibilities, I will assume, with both the modern utilitarian and the modern prioritarian, that some solution to this problem is available.

betterness-for-the-individual under conditions of risk (i.e., decision theory). It is, on reflection, distinctly odd to think that one has intuitions of ‘priority to the worse off’ form directly about such a technical notion. The modern prioritarian, however, insists that the intuitive force of her original statement remains even if ‘same size benefit’ is stipulated to mean: same size increase in this technical quantity; this is why, if making any axiological claim at all, she implicitly or otherwise advocates the value function (4).⁵

2.3 Technical Prioritarianism and Harsanyi’s aggregation theorem

The value function (4) cannot be the prioritarian’s full story: if the theory is to have any implications for real-world decision-making, (4) must be extended to an overall value function on prospects, giving the prioritarian’s judgments about which prospects are *ex ante* better (overall) than which others.

Any such prioritarian evaluation of prospects, however, will have some very odd implications. Many of these implications are easily predictable from Harsanyi’s well-known (1955) aggregation theorem, which we briefly review.

The theorem relies on the

Ex Ante Pareto principle: Let A and B be any prospects. If $A \succ_i B$ for each person i , then $A \succ B$.

Harsanyi’s theorem proves that if the individual betterness relations \succ_i and the overall betterness ordering \succ of prospects all obey the axioms of expected utility theory, and if in addition the Ex Ante Pareto principle is satisfied, then the overall betterness ordering is correctly represented by the utilitarian value function

$$V^u(\cdot) = \sum_{ij} p_j u_i^{VNM}(j \wedge \cdot). \quad (5)$$

It is clear that no prioritarian value function on prospects can satisfy this conclusion, since (4) disagrees with (5) even on the ordering of riskless prospects (that

⁵In defence of my claim that modern prioritarianism has become this ‘Technical’ claim rather than the ‘Primitivist’ version, note e.g. that the criticism of prioritarianism made in Otsuka and Voorhoeve’s (2009) — based on the ‘moral shift’ (the appearance of the transform f) that prioritarians claim to occur between first-person (individual-betterness) and third-person (overall-betterness) judgments — very obviously applies only to Technical Prioritarianism, not to Primitivist Prioritarianism. Meanwhile, of the six papers (four defending prioritarianism) that appear in a special issue of *Utilitas* (September 2012) discussing Otsuka and Voorhoeve’s criticism, none mentions the possibility that the criticism misses its mark for this reason. A Technical rather than Primitivist theory is explicitly recommended to prioritarrians in (Rabinowicz, 2002), in response to the criticisms from Broome cited above. In contrast, in much of the older work (e.g. (Nagel, 1979), (Parfit, 1997)) it is less clear whether it is an Primitivist or a Technical claim that is intended. But sometimes it is clear that the Technical version is *not* intended: thus Sen (1973), defending ‘social welfare functions’ that exhibit ‘non-linearity’ of the sort represented by the transform f in (2), writes that it would be ‘grotesque’ ‘to define a non-linear social welfare function *on von Neumann-Morgenstern utilities*’ (Sen, 1976, p. 250; emphasis in original). Similarly, when Broome (1991) argues against ‘prioritarianism’, he is certainly discussing Primitivist Prioritarianism, and does not countenance Technical Prioritarianism.

is, states of affairs). It follows that any version of Technical Prioritarianism will violate one or more of the assumptions of Harsanyi’s theorem.

Which assumption is violated depends on precisely *how* (4) is extended to prospects, and there are several ways to effect the extension. By way of salient example, the route favoured by most prioritaricians is ‘ex post prioritarianism’, expressed by the value function

$$V^{EPP}(\cdot) = \sum_{ij} p_j f(u_i^{VNM}(j \wedge \cdot)). \quad (6)$$

This theory violates the Ex Ante Pareto principle: one can easily find pairs of prospects X, Y such that X is better than Y for all affected individuals, but such that, according to ex post prioritarianism, Y is better *simpliciter* than X. Other extensions of (4) to prospects respect the Ex Ante Pareto principle, but instead violate expected utility theory (see Appendix A for the details). Any such result is very odd, and should give Technical Prioritaricians serious pause.

It is worth noting that *egalitarians* have a coherent rationale for denying the Ex Ante Pareto principle in many cases. For egalitarians⁶ think that *ex post* interpersonal inequality has intrinsic disvalue; and it can easily happen that one prospect has higher expected total von Neumann-Morgenstern utility than another, while nevertheless the second is guaranteed to exhibit a higher degree of *ex post* interpersonal equality than the first. Thus an egalitarian might well think that $A \succ B$ in the following case, despite the fact that $B \succ_i A$ for both persons involved:

Example 1: Ex post egalitarians should reject the Ex Ante Pareto principle.

	Prospect A		Prospect B	
	H	T	H	T
u_1^{VNM}	0	9	0	10
u_2^{VNM}	0	9	10	0

This rationale for rejecting Ex Ante Pareto is not, however, available to Ex Post Prioritaricians. As shown in Appendix A, V^{EPP} violates that principle even in the absence of interpersonal inequality (Example 5), and indeed even in cases involving only one person (Example 4).

Much of the recent prioritarian literature is concerned with justifying this quite general violation of the Ex Ante Pareto principle. For instance, Parfit argues that it is fully appropriate for the overall betterness relation \succ to rank some prospects involving risk lower than the individual betterness orderings \succ_i rank them (relative to prospects involving less risk), on the grounds that “[w]hen we have to make some decision on someone else’s behalf, and we don’t

⁶By definition. Here I follow Parfit’s (1997) suggestion for regimenting the terms ‘priority’ and ‘equality’. The corresponding formal issue is whether or not the value function exhibits additive separability of persons (Broome, 1991, ch. 9).

know how this person would prefer us to act, we may believe that we ought to be cautious, or risk averse” (2012, p.423). In a different vein, it might be argued that improvements in utility low down the utility scale have the status of increasing satisfaction of needs, whereas those higher up the scale have more the status of increasing facilitation of personal projects; and that it is appropriate for a third party — a “morally motivated stranger” — to give greater relative emphasis to the person’s needs over projects than it is appropriate for the person himself to give. This line of thought is offered to the would-be prioritarian by Otsuka and Voorhoeve (2009, pp. 190–2), drawing on (Scanlon, 1975, pp. 659–60) and (Nagel, 1986, pp. 166–70). An alternative tack is simply to deny that prioritarianism is in fact committed to the odd implications that examples such as the above seem to bring out, by claiming that the cases involved in those examples lie outside the domain of applicability of prioritarianism (Porter, 2012). The hope would be that a restricted domain of prospects can be found, within which prioritarianism has no unwelcome consequences.

More generally, the special issue of *Utilitas* in which the above articles by Parfit and Porter appear is devoted to replies to Otsuka and Voorhoeve’s (2009), which urges that prioritarianism errs in placing the ‘moral shift’ — the point at which considerations along the lines of ‘priority to the worse off’ kick in — at the move from betterness-for-*i* to betterness-*simpliciter*, rather than at the move from betterness-*simpliciter* of intrapersonal prospects to betterness-*simpliciter* of interpersonal distributions. Because of this alleged problem for prioritarianism, Otsuka and Voorhoeve themselves advocate responding to the supposedly problematic features of utilitarianism by moving to a view that is egalitarian, rather than prioritarian.

None of these reactions to the anti-Harsanyiian features of prioritarian value functions strikes me as promising. But arguing that lies beyond the scope of the present paper. The focus of the present paper is rather whether the initial prioritarian intuition of priority to the worse off in any case provides any sound motivation for anything like the Technical Prioritarian’s defining claim (4) in the first place; I will argue that it does not. To this end, we turn next to ‘antiprioritarian’ theory.

3 Technical Cautionism

Recall that the prioritarian’s basic driving intuition was of priority to the worse off: slightly more precisely, that it is more urgent or there is stronger reason to give a benefit of a fixed size to a worse off person than to a better off person. In parallel to this, the cautionist’s basic driving intuition is of *caution in the face of risk*. For example: Suppose that one’s fortune depends on the toss of a fair coin, and that, under the status quo, one stands to be destitute if the coin lands tails but a millionaire if the coin lands heads. Suppose further that one has the option of modifying this status quo in either (but not both) of two ways. One can arrange for some improvement to one’s condition that will materialize iff one ends up destitute. Or one can arrange for the same size improvement to

one's condition to materialize iff one ends up as a millionaire. The cautionist's intuition, slightly more precisely, is that it is better (or it is more urgent, or one has stronger reason) to opt for the former improvement to one's prospects than the latter in such a case. This intuition, he supposes, is in some conflict with the utilitarian's expected-utility method of evaluating prospects.

3.1 Clarifications

As in the case of prioritarianism, three clarifications are in order. First clarification: the cautionist does not merely assert ('Resource Cautionism') that faced with a choice between arranging for a given amount of concrete *resource* — ten sacks of rice, or a thousand dollars — to be delivered on condition one is destitute or on condition one is wealthy, it is better to opt for the former. This would be a choice between receiving some relatively small benefit if one turns out to be a millionaire and receiving a much larger benefit if one turns out to be destitute; the obvious fact that the latter is better is one that the utilitarian is perfectly well able to capture, again via the diminishing marginal utility of such concrete resources. To be distinctive, the cautionist thesis must therefore be that it is better (more urgent, etc.) to receive the benefit on condition one is worse off, even if the size of benefit delivered is independent of whether one chooses to receive the benefit in the worse or the already-better possible outcome.

Second clarification: just as prioritarians do not advocate *absolute* priority to the worse off, so cautionists do not advocate *absolute* caution. If one is choosing between, say, receiving one extra grain of rice on condition one turns out to be destitute or having one's cancer cured on condition one turns out to be a millionaire, it may yet be better (more urgent, etc.) to opt for the benefit to accrue on condition one is a millionaire. (That is, the cautionist principle is not to make the worst possible outcome as good as possible.) The ideology of 'size of benefit' is therefore crucial: it is only in cases in which 'size of benefit' is equal that cautionists will necessarily claim that (if probabilities are also equal) it is better to receive the benefit under the 'unlucky' outcome.

Third clarification: I focus on *axiological* cautionism, i.e. on the claim that the prospect that includes the extra benefit in the 'unlucky' scenario is *ex ante* better for the individual than the prospect that includes the extra benefit in the 'lucky' scenario. This is to be distinguished from a *merely deontological* cautionism, which claims that one *ought to choose* the more 'cautious' prospect but does not claim that that prospect is *ex ante* better for oneself than the alternative prospect; it is also to be distinguished from a *merely emotional* cautionism, which makes claims about what one *ought to feel* while contemplating the various prospects (say, more or less unease), but makes no claims about which is better or about which one ought to choose.

3.2 The cautionist's 'same size benefit' relation

What then, is the betterness-for-the-individual ranking of prospects according to cautionism?

Again, the issue is the relationship between the individual and overall betterness orderings. Again, it is most fruitful to approach the issue by way of the corresponding value functions. As a first pass, whereas the utilitarian holds that the individual value of a prospect is just the expectation value of the individual values of the states of affairs involved —

$$V_i^u(\cdot) = \sum_j p_j V_i(j \wedge \cdot)$$

— the cautionist insists that the individual value of a prospect is given instead by

$$V_i^c(\cdot) = \sum_j p_j g(V_i(j \wedge \cdot)), \quad (7)$$

where g is a strictly increasing and strictly concave function. This transform g expresses the desired caution.

Again, however, we have not yet said enough to pin down a genuine disagreement between cautionist and utilitarian, because the value functions V_i are as yet too indeterminate. The task is again to find some suitable way of making the notion of ‘same size increase in V_i ’/‘same size benefit’ determinate, and the same *prima facie* possibilities that we examined in the context of prioritarianism arise here.

First, one might understand ‘same size benefit’ in purely primitive terms. Again, though, this route suffers from the content problem: *merely* asserting that there is a primitive notion of benefit size does not suffice to endow that notion with content.

The (Technical) prioritarian’s response to the content problem was to precisify ‘benefit size’ by appeal to von Neumann-Morgenstern utility. It is crucial to see that this route is not available to the cautionist: it is *definitive* of von Neumann-Morgenstern utility that the expectation value $\sum_j p_j u_i^{NM}(j \wedge \cdot)$ correctly represents the relation of *ex ante* betterness-for-*i* among prospects, and hence analytically false that the modified expectation value $\sum_j p_j g(u_i^{NM}(j \wedge \cdot))$ does.

As the answer that was available to the prioritarian would render cautionism analytically false, however, so the converse is also true. The cautionist can stipulate that ‘same size benefit’ means ‘same size increase in the quantity whose sum-over-people represents overall betterness for (riskless) states of affairs’. The resulting cautionist thesis then has determinate content, and (as we spell out below) does involve a genuine disagreement with the utilitarian. This is the option our cautionist will pursue.⁷

⁷Readers may be starting to suspect that my cautionist character has lost the plot. I emphasise that I am sympathetic to this judgment: *I do not advocate his theory-building procedure*. His *raison d’être* is not to propose a tenable rival doctrine, but to raise, by way of an analogy that I will argue is quite precise, equally serious concerns about the prioritarian’s theory-building method. Section 5 will examine the cautionist’s mistakes, and will undertake the comparison between prioritarian and cautionist.

In slightly more detail: It can be proved ((Debreu, 1960); for an informal exposition, see (Broome, 1991, ch. 4)) that, provided the overall betterness ordering of states of affairs is related to the individual betterness orderings in a manner that satisfies certain axioms (principally, that the former is ‘strongly separable’ with respect to the latter), there exists a set of functions $\{u_i^D\}$ from states of affairs to real numbers, such that each u_i^D ordinally represents the corresponding betterness ordering \succsim_i , and such that, in addition, the overall betterness ordering is ordinally represented by the sum of these u_i^D , i.e. by the overall value function

$$V(\cdot) = \sum_i u_i^D(\cdot). \quad (8)$$

Call the units represented by the functions u_i^D *Debreu utiles*. Debreu utiles, then, are implicitly defined by the overall betterness ordering of states of affairs, in much the way that von Neumann-Morgenstern utiles are implicitly defined by the individual betterness orderings of prospects.

With this in hand, we continue with our development of cautionism. In terms of Debreu utiles, the utilitarian’s individual value functions for prospects are given by

$$V_i^u(\cdot) = \sum_j p_j u_i^D(j \wedge \cdot)$$

— that is, utilitarianism recommends *risk neutrality* with respect to Debreu utiles to the individual facing risk. (This follows from the fact that the utilitarian advocates (3), together with the implicit definition (8) of Debreu utiles.) But surely, holds the cautionist intuition, some risk aversion is appropriate? To capture this intuition, the cautionist insists, the individual value functions for risky prospects must be given instead by the more risk-averse formula

$$V_i^c(\cdot) = \sum_j p_j g(u_i^D(j \wedge \cdot)). \quad (9)$$

Call any position that takes the value function (9) correctly to represent the individual betterness relation among prospects *Technical Cautionism*. It is technical in the sense that its central notion of Debreu utility is not any primitive notion of quantity of well-being, but is implicitly defined by the theory of overall betterness for states of affairs. The Technical Cautionist, however, insists (misguidedly or otherwise!) that the intuitive force of his original statement remains even if ‘same size benefit’ is stipulated to mean: same size increase in this technical quantity; this is why he advocates the value function (9).

3.3 Technical Cautionism and Harsanyi’s aggregation theorem

We have noted (section 2.3) that any position that disagrees with the utilitarian on the relationship between individual and overall betterness must violate one

or more of the assumptions of Harsanyi’s aggregation theorem. This applies to Technical Cautionism no less than to Technical Prioritarianism; again, the most obvious ways of extending the claims that the Technical Cautionist has so far made so as to commit also to an overall value function for prospects results in a view that violates the Ex Ante Pareto principle. Again, this should give the Technical Cautionist pause. The various possible extensions and their respective ‘un-Harsanyian’ consequences are explored in Appendix B.

One might seek a rationale for these deviations from the Ex Ante Pareto principle and/or expected utility theory. I will not undertake that project. For what it’s worth, I suspect that anything the prioritarian might say in defence of the particular ways in which his theory violates those principles (cf. section 2.3) will have a precise analogue here, but the prospects for an ultimately successful defence strike me as dim in both cases. But the important point for present purposes is one of *symmetry* between the two positions: for all we’ve seen so far (at any rate), one should not dismiss Technical Cautionism out of hand on the basis of violations of Harsanyi’s axioms unless one also dismisses Technical Prioritarianism on the analogous basis. Let us suppose, purely for the sake of argument, that these reasons do not render either position unacceptable.

4 Prioritarianism and cautionism contrasted

Now that we have both (Technical) Prioritarianism and (Technical) Cautionism on the table, and setting aside their respective Harsanyian woes, we turn to comparing them.

In fact they are mutually contradictory. This is easy to see. We have seen that under conditions of certainty, any Technical Prioritarian subscribes to the value function $\sum_i f(u_i^{VNM})$, with f concave. Under the same conditions, the Technical Cautionist subscribes to the value function $\sum_i g^{-1}(u_i^{VNM})$, with g^{-1} convex (since g is concave).⁸ Thus, *insofar as* Technical Prioritarianism deserves the label ‘prioritarian’, Technical Cautionism is ‘antiprioritarian’; by the same token, *insofar as* Technical Cautionism deserves the label ‘cautionist’, Technical Prioritarianism is ‘anticautonist’.

We elaborate on these two last claims in turn. To see the sense in which Technical Cautionism is ‘antiprioritarian’, consider, for instance, the following case. In this example, whereas Technical Prioritarians hold that $C \succ D$, Technical Cautionists insist that $D \succ C$:

Example 2: Technical Cautionism is ‘antiprioritarian’.

⁸To elaborate: the Technical Cautionist is committed to the value function $\sum_i g^{-1}(u_i^{VNM})$ because his defining claim, *qua* Technical Cautionist, is that $u_i^{VNM} = g(u_i^D)$. It follows from this that $u_i^D = g^{-1}(u_i^{VNM})$; but all agree that the overall value function is $\sum_i u_i^D$, by definition of Debreu utility.

	Status Quo		Distribution C		Distribution D	
	Bill	Ben	Bill	Ben	Bill	Ben
u^{VNM}	2	4	3	4	2	5
u^D (Technical Prioritarian)	$\sqrt{2}$	$\sqrt{4}$	$\sqrt{3}$	$\sqrt{4}$	$\sqrt{2}$	$\sqrt{5}$
u^D (Technical Cautionist)	4	16	9	16	4	25

We stipulate the case as follows: under the status quo distribution, Bill stands to enjoy 2 von Neumann-Morgenstern utiles, while Ben will have 4. We consider giving an extra von Neumann-Morgenstern utile either to Bill (Distribution C), or to Ben (Distribution D). We then ask which distribution, C or D, is better (*simpliciter*).

Our stipulation forces all parties, including all prioritarians and all cautionists, to agree on the description of the case in terms of von Neumann-Morgenstern utiles (although nothing forces any two evaluators to agree on which concrete cases satisfy this description). Technical Prioritarians and Technical Cautionists disagree with one another, however, on the question of which description in terms of Debreu utiles is equivalent. Technical Prioritarianism's equivalent Debreu-utile description is given in the penultimate row of the above table, Technical Cautionism's in the last row. By the implicit definition of Debreu utiles, each theorist holds one distribution to be better than a second one iff the first has a higher total number of Debreu utiles than the second.

According to Technical Prioritarianism, we know, C is better than D. This can be verified by summing the Debreu-utile figures for each distribution in the penultimate row; it also seems to make intuitive sense, since Bill will be worse off than Ben even after any such increase.

According to Technical Cautionism, however, von Neumann-Morgenstern utility is a concave function of Debreu utility (for instance, the square root function). It follows that Debreu utility is a convex function of von Neumann-Morgenstern utility (for instance, the square function). Crunching the numbers in the example above, we see that according to cautionism the outcome of giving the extra von Neumann-Morgenstern utile to Ben (Distribution D) is better than that of giving the extra utile to Bill (Distribution C): according to Technical Cautionism, the former contains a total of 29 Debreu utiles, the latter only 25.

One might think that this shows Technical Cautionism to be implausible: isn't the Technical Cautionist 'giving priority' to those who are already better off? This appears to be the case, however, *only when the fundamental description of the case is taken to be in terms of von Neumann-Morgenstern, rather than Debreu, utiles*; the language of von Neumann-Morgenstern utiles biases intuitions towards Technical Prioritarianism. In the language of Debreu utiles, there is an apparently comprehensible rationale for the Technical Cautionist's judgment that it is better *simpliciter* to give the additional benefit to Ben in this case: the choice is between giving 9 Debreu utiles to the better-off Ben (who already has 16), and giving 5 Debreu utiles to the worse-off Bill (who currently has only 4). That is, if 'size of benefit' is measured in the coin of Debreu utility, the potential benefit to Bill is significantly larger than the potential benefit to

Ben. Given that all parties to the present discussion eschew *absolute* priority, it is not particularly surprising that this difference in benefit size manages to outweigh the fact that Ben is already better off.

The Technical Prioritarian, in particular, must accept reasoning of the general kind involved in the preceding paragraph, since precisely the same sort of reasoning is required in order to explain why the equally real sense in which ‘Technical Prioritarianism is anticautionist’ does not obviously render Technical Prioritarianism implausible. In the following example, whereas Technical Cautionism holds that Prospect E is better than Prospect F for the individual concerned ($E \succ_i F$), Technical Prioritarianism holds that $F \succ_i E$:

Example 3: Technical Prioritarianism is ‘anticautionist’.

	Status Quo		Prospect E		Prospect F	
	H	T	H	T	H	T
u^D	2	4	3	4	2	5
u^{VNM} (Technical Cautionism)	$\sqrt{2}$	$\sqrt{4}$	$\sqrt{3}$	$\sqrt{4}$	$\sqrt{2}$	$\sqrt{5}$
u^{VNM} (Technical Prioritarianism)	4	16	9	16	4	25

Example 3 involves a single flip of a fair coin, and only one person, Claire (say). We stipulate the case as follows: in the status quo, Claire stands to enjoy 2 Debreu utiles if the coin lands Heads, 4 if it lands Tails. We consider modifying the prospect to increase Claire’s well-being by one Debreu utile either just in case the coin lands Heads (Prospect E), or just in case the coin lands Tails (Prospect F). We then ask which modified prospect, E or F, is (*ex ante*) better for Claire.

In precise analogy to the above: our stipulation forces all parties, including all prioritariums and all cautionists, to agree on the description of the case in terms of Debreu utiles (again, they may not agree which concrete cases satisfy this description). Technical Cautionism and Technical Prioritarianism disagree on the question of which description in terms of von Neumann-Morgenstern utiles is equivalent. By the implicit definition of von Neumann-Morgenstern utiles, each theorist holds one prospect to be better for Claire than a second one iff the first has a higher expected number of von Neumann-Morgenstern utiles than the second.

According to Technical Cautionism, we know, Prospect E is better for Claire than Prospect F. This can be verified by calculating Claire’s expected number of von Neumann-Morgenstern utiles under each prospect, according to the penultimate row of the above table. It also makes intuitive sense for the prospect that improves Claire’s lot under the Heads outcome (rather than Tails) to be better for Claire, since she will still be worse off if the coin lands Heads than if it lands Tails even after any such improvement.

According to Technical Prioritarianism, however, Debreu utility is a concave function of von Neumann-Morgenstern utility (in our example, the square root function). It follows that von Neumann-Morgenstern utility is a convex function

of Debreu utility (here the square function). As a result, in Example 3, according to Technical Prioritarianism, improving Claire’s lot in the Tails state of nature turns out to result in a prospect that is *ex ante* better for Claire, notwithstanding the fact that she *already* stood to be better off under Tails than under Heads: according to Technical Prioritarianism the former option results in an expected von Neumann-Morgenstern utility of $\frac{25}{2}$, the latter $\frac{29}{2}$.

It might be suggested that this shows Technical Prioritarianism to be implausible: isn’t the Technical Prioritarian implausibly recommending to the individual the very opposite of caution in the face of uncertainty? How can the judgment that it is a greater improvement in Claire’s prospects for her to receive the extra Debreu utile under the state of nature in which she is already better off, rather than improving her lot in case the coin lands Heads, possibly be rationalised? As above, the Technical Prioritarian has an answer to this charge, but it is an answer that depends on shifting to her preferred language. Technical Prioritarianism seems implausibly anticautious *only when the fundamental description of the case is taken to be in terms of Debreu, rather than von Neumann-Morgenstern, utiles*; the language of Debreu utiles biases intuitions towards Technical Cautionism. In the language of von Neumann-Morgenstern utiles, there is an apparently comprehensible rationale for the Technical Prioritarian’s judgment that Claire’s prospects are improved more by securing the additional Debreu utile in case the coin lands Tails: the choice is between gaining 9 von Neumann-Morgenstern utiles if the coin lands Tails (in which case she already has 16) , and gaining 5 von Neumann-Morgenstern utiles if the coin lands Heads (in which case she currently has only 4). That is, if ‘size of benefit’ is measured in the coin of von Neumann-Morgenstern utility, the potential benefit in the Tails eventuality is significantly larger than the potential benefit in the Heads eventuality. Given that all parties to the present discussion eschew *absolute* caution, it is not particularly surprising that this difference in benefit size manages to outweigh the fact that Claire already stands to be better off under Tails than under Heads.

To sum up this section: Technical Prioritarianism and Technical Cautionism are mutually contradictory. Because of this, the intuition driving each can be used to make the other appear *prima facie* implausible. In each case, however, the defendant is able to give a comprehensible rationale for his apparent violation of the other party’s driving intuition in terms of benefit size, *provided* he is permitted to tie ‘benefit size’ talk to his preferred cardinalisation of the well-being scale (von Neumann-Morgenstern or Debreu utiles respectively). In every aspect that I have discussed, the situation between the two theories is symmetrical.

5 Basic intuitions and Technical axiologies

We have seen (section 4) that Technical Prioritarianism and Technical Cautionism contradict one another. Therefore, they cannot both be true. On the other hand, the basic intuitions that are supposed to be behind each, that it is appro-

appropriate to give priority to the worse off in distributive matters, and that caution is appropriate in the face of risk, clearly do (for all their vagueness) each contain some important elements of truth. Therefore, in at least one case, the driving intuition contains some important elements of truth while the theory that is supposed to capture it is false. The present section investigates the implications of this for the motivation of Technical Prioritarianism.

5.1 The basic intuition of caution does not support Technical Cautionism

Let us suppose (as we must, if we are to be as sympathetic as possible to the Technical Prioritarian) that the Technical Cautionist takes a mis-step: let us suppose, that is, that while the basic intuition of caution in the face of risk has a lot going for it, the Technical Cautionist axiological theory is false. We then require an explanation of where the journey from intuitions of caution to Technical Cautionism went wrong.

Some readers, no doubt, will feel impatience with this task: it is, they clearly see, *obvious* that Technical Cautionism is not the right way to capture the elements of truth in the basic intuition of the appropriateness of caution, and, furthermore, nothing has been said (in the present paper or otherwise) by way of argument to the contrary. I agree; sometimes, however, there is significant illumination to be gained from a precise diagnosis of the nature of the mistake even in the obviously mistaken. My claim in the present case is that once we have seen what is mistaken in each of the possible routes to Technical Cautionism, we should come to think that the same mistake is being made in the available routes from the intuition of priority to the worse off to Technical Prioritarianism. Only by making the precise nature of the mistakes in question clear can we be in a position to assess whether or not this is the case.

We must ask, then, what lines of thought *might conceivably* lead a theorist from the basic idea of the appropriateness of caution to the Technical Cautionist theory. There are two.

The first begins from the basic (and vague) intuition of the appropriateness of caution, acknowledges that it is as it stands pretty vague, and adds the thought that Resource Cautionism does not seem to capture all that we have in mind when we say that caution is appropriate in the face of risk — it does not, that is, seem to amount to a fully satisfactory precisification of the initial intuition. The argument in this case might be:

- (P1) The elements of truth in the basic intuition of caution in the face of risk are not exhausted by the claims of Resource Cautionism: to rest content with that theory is not to take the intuition of caution seriously enough.
- (P2) Primitivist Cautionism fails to have any determinate content (the content problem).
- (P3) An axiology recommending *absolute* caution is implausible.

- (P4) One other theory that does more justice to the intuition of the appropriateness of caution than either Resource or Primitivist Cautionism is Technical Cautionism (since the latter theory holds that ‘risk aversion’ is appropriate not only with respect to physical resources or any primitive notion of well-being, but even with respect to Debreu utility).
- (P5) There is no other way, besides Absolute Cautionism, Primitivist Cautionism and Technical Cautionism, of doing justice to the sense that Resource Cautionism does not fully capture the appropriateness of caution.

Therefore,

- (C) Technical Cautionism is true.

This is not a good argument. (P3) is uncontroversial (in the present setting), and I have argued for (P2). It is not *obvious* that (P1) is true, but let us grant it for the sake of the argument; there are more glaring errors here.

The most obvious one is that (P5) is false. There are *non-axiological* facts capturing important parts of the intuitive sense of the appropriateness of caution, which are necessarily not captured by any purely axiological theory (such as Resource Cautionism). Most obviously, part of what is going on when one insists on the appropriateness of caution in the face of risk is an assertion about the structure of appropriate deliberation. It is appropriate *carefully to consider* whether one is taking the right decisions in the face of risk, and to include in one’s deliberations attention to the possibility (and estimates of the probability) of unfavourable outcomes. It is also appropriate to feel a *sense of unease* when taking risks involving significant probabilities of significantly worse outcomes (emotional cautionism). Furthermore, there are complex *deontological* aspects of the situation when questions of risk-taking on behalf of others arise that arguably go beyond axiology, arising from the value of autonomy in conjunction with the possibility of divergent estimates of the probabilities and values involved.⁹ These, I think, are the directions to look for completion of the picture, if one is troubled by the thought expressed in (P1). One mistake made by the Technical Cautionist who follows the path of our first argument, then, is a neglect of the possibility of non-axiological aspects of the situation, leading him to seek more in the purely axiological aspect of the story about caution than can plausibly be found there. Once he takes full account of the ability of non-axiological aspects to add to the axiological picture, he may lose his sense that the Resource Cautionist axiology does not go far enough even *qua* axiology.

It is also far from clear that (P4) is true. Technical Cautionism does ‘go beyond’ Resource and Primitivist Cautionism in the sense that it is not a sub-theory of either of the first two — it makes a claim that neither Resource nor Primitivist Cautionism does — but it is far from clear that this additional claim amounts to *doing more justice to the initial intuition of caution*. This point will

⁹This is where the issues that Parfit touches on in the comments on p.423 of his (2012) (quoted in section 2.3 above) fit into the discussion; *pace* Parfit, they are not relevant to the question of *axiology*.

become clearer in discussion of the second possible argument for Technical Cautionism, to which we now turn.

The second ‘argument’ does not in fact have enough structure to deserve that honorific. Rather than starting from an intuition of the appropriateness of caution that is admitted to be vague and arguing about what does or does not fully capture it, the second line of thought simply claims direct intuitive support for a far more precise claim, viz. that relative to any ‘status quo’ prospect for a given individual that involves two states of equal probability, the prospect that results from increasing the individual’s well-being by a given number of Debreu utiles in the state of nature in which she is worse off is *ex ante* better for her than the prospect that results from increasing the individual’s well-being by the same number of Debreu utiles in the state of nature in which she is better off. Further, it claims that this is a manifestation of caution in the face of risk. But the claim in question just *is* Technical Cautionism. So, our second Technical Cautionist claims, the theory itself simply has direct intuitive support, as part of the intuition of caution in the face of risk.

The correct reply to our second Technical Cautionist character, I think, is not merely announcement that we do not share the intuition in question, but also profound scepticism about whether he really has the intuition that he takes himself to have. For, in the first instance, when we unpack the statement that he claims to have a direct intuition about, it does not look at all like the sort of statement one would expect to have unstructured intuitive access to; in particular, it is not plausibly taken as an endorsement of caution in the intuitive sense. Secondly, an error theory is available, to explain why our second Technical Cautionist might *think* he has this intuition, when in fact he does not, and why he might *think* that his statement expressed caution in the intuitive sense, when in fact it does not.

Taking the first aspect first: The claim in question involves a technical term, ‘Debreu utility’. That technical notion has a stipulative definition. Therefore, any intuition about the truth of a proposition that can be expressed using this term must *ipso facto* be an intuition about the truth of a proposition that is also expressible by substituting its definition. In the present case, the (inevitably cumbersome) sentence resulting from such substitution is: “Relative to any ‘status quo’ prospect for a given individual, involving two states of equal probability, the prospect that results from increasing the individual’s well-being by a given number of units of the quantity whose sum-over-people correctly represents betterness-*simpliciter* in the absence of risk in the state of nature in which she is worse off is *ex ante* better for her than the prospect that results from increasing the individual’s well-being by the same number of units of the quantity whose sum-over-people correctly represents betterness-*simpliciter* in the absence of risk in the state of nature in which she is better off.” Spelling out the full content of the claim in this way, though, highlights the fact that the claim being expressed is not a monadic one about the betterness-for-the-individual relation, but rather a *comparative* one, about the *relationship between* the betterness-for-the-individual relation and the betterness-*simpliciter* relation. As such, it is not an expression of the statement that caution is appropriate in the face of risk:

it is, rather, the statement that the *degree* of caution (with respect to physical resources, or primitive well-being, or any other independently fixed scale) that figures in the individual-betterness relation in the face of risk is *greater than* the *degree* of inequality-aversion (with respect to the same scale) that figures in the betterness-*simpliciter* relation in the absence of risk. It is then far less clear that one can have any direct intuition about this more complex sentence. One needs, rather, carefully to think through the implications before deciding on which side intuition falls. In particular, one needs to notice that it is equivalent to asserting that degree of inequality aversion is *less* than degree of risk aversion (in the same two senses), and that equivalent assertion is perhaps intuitively less comfortable. One also needs to notice the Harsanyian troubles that follow from it, discussed briefly in section 3.3 and elaborated in Appendix B. I conclude that it is not plausible that our Technical Cautionist really has an intuition in support of the claim in question, and that in any case that claim is not an expression of caution.

Moving on to the second aspect: the error theory is as follows. Our character has internalised the point that axiological representations of the notion of caution almost invariably amount to claims that the quantity whose expectation value represents *ex ante* betterness for the individual (that is, by definition, von Neumann-Morgenstern utility) is a concave transform of some independently identified quantity. He then makes the mistake of thinking that taking the notion of caution *really seriously* amounts to accepting such a concavity claim about *any* ordinal representation of individual well-being you care to name – the details (he assumes) matter not. As a result, when offered the statement “von Neumann-Morgenstern utility is a concave transform of Debreu utility”, he does not pause to contemplate the meaning of ‘Debreu utility’, beyond checking that ordinally it represents individual well-being (i.e., that if one outcome is better-for-the-individual than another, the first has higher Debreu utility than the second); he simply accepts that since this statement has the general form of a cautionist claim, his intuition supports it. On reflection, it is obvious that this procedure is unreliable, since even an unrepentant Technical Cautionist agrees that there are *some* quantities that ordinally represent individual well-being and with respect to which risk aversion is *not* appropriate (for example, von Neumann-Morgenstern utility itself, square root of von Neumann-Morgenstern utility, and perhaps cube root of Debreu utility). According to the error theory, then, when our second Technical Cautionist is consulting his intuitions on the sentence in question, he is failing to grasp the proposition that is in fact expressed by that sentence. He does not really have any direct intuition in support of that proposition.

To sum up our findings, then: a Technical Cautionist who takes his theory to be motivated by the basic intuition of caution is making a mistake. Part of the reason for his wrong turn might be a neglect of the possibility of non-axiological aspects of the full account of caution, leading him to seek more in the purely axiological aspect than can be found there. Another part is his failure to notice that any claim about the prudential value of Debreu utility is not a claim about the prudential value of well-being understood intuitively, but

rather a comparative claim about the relationship between prudential value and moral value: he thereby also fails to notice that assessment of such claims must pay attention to the details of the stipulative definition of Debreu utility, and must not uncritically assume that intuitions about “well-being” apply equally to Debreu utiles.

5.2 The basic intuition of priority to the worse off does not support Technical Prioritarianism

So much for Technical Cautionism. It has now served its dialectical purpose; we can consign it for evermore to the scrap-heap on which it belongs. For methodological consistency, though, if we tell the above story about Technical Cautionism, and if in addition we cannot point to any relevant asymmetry between the cautionist story and the prioritarian one, then we must tell the same story about Technical Prioritarianism.

That story about Technical Prioritarianism is as follows. There are two lines of thought that might conceivably lead a theorist from the basic idea of the appropriateness of priority to the worse off to the Technical Prioritarian theory. The first begins from the basic (and vague) intuition of priority to the worse off, acknowledges that it is as it stands pretty vague, and adds the thought that Resource Prioritarianism does not seem to capture all that we have in mind when we say, e.g., that it is more urgent to benefit the worse off. As above, this line of thought can be turned into an argument for Technical Prioritarianism *if* one assumes, crucially, that there are no possibilities for capturing the elements of truth in the basic intuition of priority other than the Resource Prioritarian, Primitivist Prioritarian, Absolutist Prioritarian and Technical Prioritarian axiologies (or axiology-attempts). (The argument precisely parallels that given from (P1)–(P5) above for Technical Cautionism.) As above, though, such a line of argument would be mistaken, at least in its neglect of the non-axiological elements in the initial prioritarian intuition. No purely axiological claim, for instance, will capture the facts about fitting feelings (for instance, that it is fitting to feel a greater sense both of urgency and of empathetic distress when contemplating aspects of distributive problems involving the relatively badly off than those involving the relatively well off, quite aside from one’s decisions on which distribution is in the end optimal). It is therefore inevitable that no axiology will itself provide full satisfaction to one seeking to capture every aspect of his intuitions of priority.

The second line of thought, rather than starting from an intuition of priority to the worse off that is admitted to be vague, simply claims direct intuitive support for a far more precise claim, viz. that relative to any ‘status quo’, the state of affairs that results from increasing a worse-off individual’s well-being by a given number of von Neumann-Morgenstern utiles is better than the state of affairs that results from increasing a better-off individual’s well-being by the same number of von Neumann-Morgenstern utiles. Further, it claims that this is a manifestation of priority to the worse off. But the claim in question just is Technical Prioritarianism. So, our second Technical Prioritarian claims,

the theory itself simply has direct intuitive support, as part of the intuition of priority to the worse off.

I submit that the correct reply to this second Technical Prioritarian is profound scepticism about whether she really has the intuition she takes herself to have. Precisely as in the case of the analogous Technical Cautionist, unpacking the statement that she claims to have a direct intuition about reveals it to be a statement that one would not expect to have unstructured direct intuitive access to, and in any case it is not plausibly taken as an endorsement of ‘priority to the worse off’ in the intuitive sense. And, also as in the cautionist case, error theories are available to explain the prioritarian’s error.

I expect these claims to meet with far more resistance than the analogous claims about Technical Cautionism, so, while the following is easily predictable from the above treatment of Technical Cautionism, I must again ask for the reader’s patience: I will set out the target of such resistance explicitly, to provide the resistance with something to shoot at.

Our second Technical Prioritarian’s claim involves a technical notion, viz. that of von Neumann-Morgenstern utility. *That technical notion has a stipulative definition.* (It is easy to forget this, because the notion of von Neumann-Morgenstern utility is by now comparatively familiar. It is therefore relatively natural, psychologically speaking, to assume that von Neumann-Morgenstern utility just is the intuitive notion of quantity of well-being. But *that assumption is false*; we must stick to the definition.) Therefore, any intuition about the truth of a proposition that involves this notion must *ipso facto* be an intuition about the truth of a proposition that is also specifiable by substituting its definition. In the present case, the (inevitably cumbersome) sentence resulting from such substitution is: “Relative to any status quo, the state of affairs that results from increasing a worse off individual’s well-being by a given number of units of the quantity whose expectation value correctly represents *ex ante* betterness-for-the-individual of prospects is better-*simpliciter* than the state of affairs that results from increasing a better off individual’s well-being by the same number of units of the quantity whose expectation value correctly represents *ex ante* betterness-for-the-individual of prospects”. Spelling out the full content of the claim in this way, though, again highlights the fact that the claim being expressed is not a monadic one about the betterness-*simpliciter* relation, but rather a *comparative* one, about the *relationship between* the betterness-for-the-individual and the betterness-*simpliciter* relations. As such, it is not an expression of priority to the worse off: it is, rather, the statement that the *degree* of priority to the worse off (with respect to physical resources, or primitive wellbeing, or any other independently fixed scale) that figures in the betterness-*simpliciter* relation in the absence of risk is *greater than* the *degree* of aversion to risk (with respect to the same scale) that figures in the betterness-for-the-individual relation under risk. It is then far less clear that one can have any direct intuition about this more complex sentence. One needs, rather, carefully to think through the implications before one decides on which side intuition falls. In particular, one needs to notice that it is equivalent to asserting that degree of appropriate risk aversion is less than degree of appropriate prioriti-

sation of the worse off, and that equivalent assertion is perhaps intuitively less comfortable; one needs also to notice the Harsanyian troubles that follow from it, discussed in section 2 and appendix A. I conclude that it is not plausible that our Technical Prioritarian really has a direct intuition in support of the claim in question, and that in any case that claim is not an expression of priority to the worse off.¹⁰

We have, moreover, a transparent error theory to explain why, if my claims are correct, prioritaricians tend so often to miss this. The Technical Prioritarian has internalised the point that axiological representations of the notion of priority to the worse off almost invariably amount to claims that the quantity whose sum-over-people correctly represents betterness-*simpliciter* (that is, by definition, Debreu utility) is a concave transform of some independently identified quantity. She then makes the mistake of thinking that taking the notion of priority *really seriously* amounts to accepting such a concavity claim about *any* ordinal representation of individual well-being you care to name – the details (she assumes) matter not. As a result, when offered the statement “Debreu utility is a concave transform of von Neumann-Morgenstern utility”, she does not pause to contemplate the meaning of “von Neumann-Morgenstern utility”, beyond checking that ordinarily it represents individual well-being (i.e., that if one outcome is better-for-the-individual than another, the first has higher von Neumann-Morgenstern utility than the second); she simply accepts that since this statement has the general form of a prioritarian claim, her intuition supports it.

The general psychological tendency postulated by this error theory, in both the prioritarian and the cautionist case, has been confirmed in psychological studies (Greene & Baron, 2001)¹¹. On the prioritarian side, it also fits the literature extremely well. The following passage from (Parfit, 2012) is typical. Discussing the content problem for (what I am calling) Primitivist Prioritarianism, Parfit writes:

Of the writers who have criticized the Priority View, some suggest that this view may not make sense. If we believe that, by benefiting Tom, we would do more to make the outcome better, this may show that, on our view, the benefit to Tom would be greater than the benefit to Ted. To defend the Priority View, we must explain what it would mean to claim that one of two benefits would be smaller,

¹⁰In recommending to prioritaricians that they embrace Ex Post (Technical) Prioritarianism, Rabinowicz (2002) notes that since the issues of priority to the worse off and of prudential risk aversion are conceptually entirely distinct, it is not *immediately inevitable* that these two degrees will coincide. This observation is correct: there is some argument to be had here (it is the argument over the Ex Ante Pareto principle). But we must not overstate the degree of comfort that the Technical Prioritarian can take from this observation: to note that priority of the worse off and prudential risk aversion are conceptually distinct is not to supply any positive reason for thinking that the second is smaller than the first.

¹¹Greene and Baron’s experiment shows, for various sorts of utility, that subjects report having intuitions of declining marginal utility not only with respect to independent scales (such as years of life), but also with respect to *that same sort of utility*: intuitions whose content is in fact incoherent.

though this benefit would contribute more to the value of the outcome. These sceptics doubt that we can distinguish between what is better for people and what makes the outcome better.

... This objection raises difficult questions about the concepts *good* and *good for*, and about what it would be for some benefits to be greater than others, and for some outcomes to be better. Though I believe that this objection can be answered, I shall not try to show that here. *I shall simply assume that we can distinguish between the size of different possible benefits and the amount by which these benefits would make the outcome better, so that we can intelligibly reject Utilitarian beliefs.* (*ibid.*, p.403; emphasis added)

So far, so good; but Parfit then proceeds, in the remainder of the paper cited, uncritically to use the language of ‘benefit’ and ‘greater benefit’, indexing this to his unspecified cardinalisation of the well-being scale (having stipulated only that his cardinalisation is other than the Debreu one), and taking prioritariums to have intuitions in favour of prioritarian-style claims that are robust despite that underdetermination. In particular, the von Neumann-Morgenstern cardinalisation is as good as any other according to this stipulation, so he takes his defence to be a defence of Technical Prioritarianism.

This form of argument is necessarily unreliable. On reflection, this is obvious, since even an unrepentant Technical Prioritarian agrees that there are *some* quantities that ordinarily represent individual well-being and but with respect to which a concave relationship to Debreu utility is not appropriate (for example, the Debreu utility itself, the square root of Debreu utility, and perhaps the cube root of von Neumann utility). According to the error theory, then, when our second Technical Prioritarian is consulting her intuitions on the sentence in question, she is failing to grasp the proposition that is in fact expressed by that sentence. She does not really have any direct intuition in support of that proposition.

It will do the Technical Prioritarian no good to object that von Neumann-Morgenstern utiles *really* correspond to quantity of well-being or ‘benefit size’, while Debreu utiles are indeed a gerrymandered quantity that it is far harder to have any intuition about. This claim is just not plausible. *Both* notions, that of von Neumann-Morgenstern utility and that of Debreu utility, are theoretical; each issues from a representation theorem that is concerned with the representation of some evaluative ordering, and a similar level of complexity of reasoning is involved in the two theorems. It is only a combination of sociological accident (the Debreu representation theorem being less well-known than those of decision theory) and interest (seeking a ‘Technical’ theory that encodes the intuition of priority to the worse off, rather than one that encodes the intuition of caution in the face of risk — as noted above, this rules out the Debreu scale on pain of inconsistency) that explains why so many authors have been happy uncritically to accept the von Neumann-Morgenstern utility scale as the one their quantity-of-well-being intuitions are ‘really about’.

I have drawn two conclusions: that the Technical Prioritarian does not really

have any direct intuition in favour of her central claim, and that in any case that claim is not an expression of priority to the worse off. I am more confident of the second conclusion than the first. There certainly will be Technical Prioritarians who insist, even after the content of the claim in question has been fully spelled out (i.e. the language of von Neumann-Morgenstern utility substituted away, as above), that they have a robust intuition in favour of that claim. These are the theorists discussed on page 10, who believe that there is some *independent* reason to think that the relation between the betterness-for-the-individual and the betterness-*simpliciter* relation is as the Technical Prioritarian says (for example, a reason relating to the nature of appropriate choice on behalf of others whose preferences are not known, or relating to the differences between the attitude-profiles of an agent versus a morally motivated stranger to that agent's needs on the one hand, and personal projects on the other). As I mention there, while I do not think that any such reason is convincing, I have not argued for that in the present paper. For all the present paper has said, therefore, there may be some intuitive support for Technical Prioritarianism after all. There is no point in trying to ascertain whether intuitive support of the character in question would count as 'direct', and therefore whether my first conclusion would still stand in such a case. The important conclusion is the second: any such reason for agreeing with the Technical Prioritarian on the relationship between betterness-for-the-individual and betterness-*simpliciter* would not be a manifestation of the appropriateness of priority to the worse off.

The reason this is important is that failure to recognise it has two adverse effects, especially in the context of Harsanyi's (1955), Otsuka and Voorhoeve's (2009), and their respective aftermaths. Firstly: if there was a good argument from 'priority to the worse off' to Technical Prioritarianism, then, in evaluating the case for and against Technical Prioritarianism, on the 'for' side we would combine the weight of arguments from 'priority to the worse off' with independent reasons for thinking that individual and overall betterness diverge in the way the Technical Prioritarian postulates. In that case, Technical Prioritarianism might well emerge victorious even if the *independent* reasons in its favour are much weaker than the Harsanyi-inspired case against it. If I am correct, on the other hand, then the intuition of priority to the worse off has nothing to do with Technical Prioritarianism, and those independent reasons must stand on their own against the Harsanyi-style 'against' case. And they are much less likely to win the latter battle.

Secondly, and relatedly: There is good news for the basic prioritarian intuition here. If (as I have argued) Technical Prioritarianism is not any precisification of that intuition, then while it does remain of interest to assess the extent to which the points of Harsanyi, and Otsuka and Voorhoeve, damage the axiology that I have been calling Technical Prioritarianism, the basic prioritarian intuition is not at stake in any such assessment. Prioritarians in the original, intuitive, sense need not be on the defensive. If 'a prioritarian theory' means: a theory that is motivated by the basic intuition of priority to the worse off, then *Technical Prioritarianism is not a prioritarian theory*.

6 On the motivation for Technical Egalitarianism

This paper has so far focussed on prioritarianism: that is, on positions that attempt to capture intuitions of ‘priority to the worse off’ while agreeing with utilitarianism that the overall value function, at least in the absence of risk, is additively separable. Do its arguments apply also to egalitarian positions, i.e. those subscribing to value functions that are *not* additively separable? The answer is affirmative, but the details are more complicated than in the prioritarian case, for a distracting technical reason.

To address the issue, it is helpful to isolate some element of the basic intuition that the distribution of well-being is important that is itself neutral between prioritarianism and egalitarianism. An obvious candidate is the

Pigou-Dalton Principle: Given any states of affairs S_1 and S_2 involving the same population, and any individuals A and B in this population such that A is better off than B in S_1 , if S_2 is obtained from S_1 by transferring a fixed amount of well-being from A to B , and the distance between A and B ’s well-being levels is less in S_2 than in S_1 , then B is better than A .

As noted by Adler and Sanichirico (2006), the resulting principle is satisfied not only by Technical Prioritarianism, but also by various value functions that are not additively separable, even in the absence of uncertainty — that is, by value functions that are definitely “egalitarian” rather than “prioritarian”.¹²

The Pigou-Dalton principle sounds intuitively highly plausible, especially if one supposes that there is some primitive notion of ‘amount of well-being’ to which we have intuitive access, and that that is the notion the principle is about. In modern applications of the principle, however, for the same reasons as in the discussion of prioritarianism, authors typically interpret ‘amount of well-being’ as amount of *von Neumann-Morgenstern utility*; they do not, however, pause seriously to re-assess the plausibility of the principle in light of this re-interpretation of one of its key terms.

Setting the Pigou-Dalton principle aside for a moment, let us consider the following ‘cautionist’ analogue:

Cautionist Transfer Principle: Given any lotteries L_1 and L_2 involving the same probabilities, and any equiprobable states of nature S_1 and S_2 such that in L_1 , a given individual is better off under S_1 than under S_2 , if L_2 is obtained from L_1 by transferring a fixed amount of that individual’s well-being from S_1 to S_2 , and the distance between the individual’s well-being in S_1 and S_2 is smaller in L_2 than in L_1 , then L_2 is *ex ante* better for the individual in question than L_1 .

This principle is, I submit, just as intuitively plausible as the Pigou-Dalton Transfer Principle: insofar as the Pigou-Dalton principle is an expression of

¹²An example is the value function $V = (\sum_i u_i^{VNM})(1 - G)$, where G is the ‘Gini coefficient’, a measure of inequality. See e.g. (Adler & Sanichirico, 2006, p. 301–2).

concern for distribution, the Cautionist Transfer Principle is an expression of the intuitively reasonable caution in the face of risk. And just as the original Pigou-Dalton principle is neutral between prioritarianism and egalitarianism, one might here proceed to formulate risk-averse individual value functions — representing betterness for the individual among prospects — that do not treat distinct states of nature in an additively separable manner, as well as ones that do. (Of course, any such non-additively-separable individual value function will violate one or more of the axioms of expected utility theory.) If the individual value function *does* exhibit additive separability and if ‘amount of well-being’ is here interpreted by reference to Debreu utility, the resulting (Technical Cautionist) value function will violate the original Pigou-Dalton principle (interpreted in terms of von Neumann utility). Correspondingly, of course, any additively separable theory satisfying the original Pigou-Dalton principle (interpreted in terms of von Neumann-Morgenstern utility), i.e. any Technical Prioritarian theory, will violate the Cautionist Transfer Principle (interpreted in terms of Debreu utility).

So far, so good; now for the special complication. The reason why the dialectic between the Technical *Egalitarian* and Technical Cautionist is more complicated than that between Technical Prioritarian and Technical Cautionist is that in our discussion in section 3, the additive separability of the prioritarian value function (4) enabled us to define a notion of utility (Debreu utility) that was based on the overall value function *and that even the prioritarian could have no objection to*, in analogue to the definition of von Neumann-Morgenstern utility that is underwritten by expected utility theory (and which even the Technical Cautionist has no objection to). When dealing with a (Technical Egalitarian) interlocutor whose value function is not additively separable, however, *this* definition (at any rate) will not create mutually acceptable terminology: unlike the prioritarian, the egalitarian should insist that there is no such thing as an individual’s Debreu utility. This provides him with an unhelpful ‘stonewalling’ tactic that was not available to the prioritarian: pending some analogous way of deriving a well-defined measure of ‘individual well-being’ from an overall value function that is not additively separable, the egalitarian can insist that there is no way of giving content to a would-be Technical Cautionist claim that does not rely on false presuppositions about the structure of the overall value function. It is also the reason why we could not say straightforwardly, in the preceding paragraph, that a Technical Egalitarian theory violates the Debreu-utility version of the Cautionist Transfer Principle. (The dialectical situation is, of course, precisely analogous to that between a prioritarian who formulates his claims in terms of von Neumann-Morgenstern utility, on the one hand, and a cautionist whose individual value function is not additively separable on the other; since the latter rejects expected utility theory, he does not recognise the notion of von Neumann-Morgenstern utility.)

Notwithstanding this complication, however, the Technical Egalitarian too should be perturbed by the intuitive plausibility of the Cautionist Transfer Principle. It cannot be that the cautionist intuition creates no problems for him simply for the formal reason that the overall value function is not additively

separable (a matter that is itself essentially *neutral* on the issue of inequality aversion). As in our discussion of prioritarianism, the plausibility of the Cautionist Transfer Principle highlights the importance of the choice of utility measure; and the possibility of making it precise using the notion of Debreu utility, even if that is not a notion that the egalitarian himself recognises, highlights the dangers of appealing to primitive evaluative intuitions when settling the *relationship between* two value functions (individual and overall) each of which is itself an evaluative matter.

7 Conclusions

Early prioritarians were ‘Primitivist Prioritarians’: they took their thesis to be about a primitive notion of quantity of well-being that has content independently of the representation of any evaluative ordering. Critics, however, doubted that there exists any such notion. In response to this, in the recent prioritarian literature, prioritarians have instead adopted ‘Technical Prioritarianism’. That is, they have taken the prioritarian claim to be about the technical notion of von Neumann-Morgenstern utility. This does succeed in supplying *a* means of stipulatively defining ‘same size benefit’ that gives the prioritarian’s key axiological claim determinate content, but I have argued that insofar as the prioritarian’s motivation is the intuition of ‘priority to the worse off’, this is a mistake: she has no reason to want *that* content. One can see that it is a mistake, if not otherwise, by noting that for every line of argument that might underwrite the move from the basic prioritarian intuition to Technical Prioritarianism, a precisely analogous line of argument takes one from the intuition of ‘caution in the face of risk’ to a theory (‘Technical Cautionism’) that directly contradicts Technical Prioritarianism. Technical Prioritarianism *may* be true — at any rate, I have not argued against it here — but in the light of difficulties encountered in extending it to cover cases involving risk, this is *prima facie* unlikely, and in any case the basic intuition of ‘priority to the worse off’ provides it with no motivation. It can be motivated, if at all, only by arguments that pay far more attention to the details of the definition of von Neumann-Morgenstern utility than the basic prioritarian intuition does. The Technical Prioritarian’s claim (after all!) is that it is better to give a fixed increase in the quantity whose expectation value correctly represents *ex ante* betterness for the individual under conditions of risk to a worse off individual than to a better off individual; equivalently (given shared presuppositions), that it is *ex ante* better-for-an-individual to enjoy a fixed increase in the quantity whose sum-over-persons represents betterness-*simpliciter* on condition she is already better off, than on condition she is worse off (if probabilities are equal). Any argument for Technical Prioritarianism therefore has to be an argument for this claim, in all its gory detail. Since the arguments from “priority to the worse off” appeal only to an unspecified measure of “well-being”, they do not have the required character.

I make two recommendations to those whose only motivation in this vicinity

is the intuition of priority to the worse off, and who are not satisfied with Resource Prioritarianism. First: that they look outside of axiology (and in particular to issues of fitting attitude) for some important parts of their view, lest they be deceived into seeking more than is appropriate in the axiological part. Second: that, *if* they deem Resource Prioritarianism insufficient even *qua* axiological claim, they return to the drawing board that is presented by the content problem, seek some further way of making ‘size of benefit’ precise, and investigate a value function that is prioritarian in form *relative to this further cardinalisation of well-being*. But in assessing the latter value function, the details of the cardinalisation are essential: the exercise cannot be carried out in the abstract. We should then note that none of the claims that a prioritarian will be led to by following either or both of these recommendations is incompatible with modern (i.e. ‘Technical’) utilitarianism, since the latter asserts only the coincidence of von Neumann-Morgenstern and Debreu utility. Relatedly, none will commit the prioritarian to violation of any of Harsanyi’s axioms, so the problems discussed in section 2.3 need not arise. Nor will any be in conflict with the claims that a *cautionist* would be led to by following the analogous recommendations regarding how a would-be cautionist should respond to the arguments of section 5.1: it is only *Technical* Prioritarianism and *Technical* Cautionism that are mutually incompatible, not prioritarianism and cautionism *per se*. Prioritarians following these recommendations therefore also avoid facing the task of explaining why their theory is any better-motivated than the formally analogous ‘cautionist’ one.

Parfit (2012, p.423) suggests that when utilitarians think it obvious that one should deliver the maximum expected benefit, they are misled by the term ‘expected’.¹³ If the argument of the present paper is correct, there is irony here: on the contrary, if and when Technical Prioritarians think it obvious that maximising expected benefit gives insufficient priority to the worse off, they are misled by the word ‘benefit’. Their problems arise from simultaneously agreeing to use ‘(same size) benefit’ as a technical term, and building their theory on intuitions drawn from its non-technical usage.

I have argued also that the same concerns apply to egalitarianism: basic driving intuitions about the value of equality carry no commitment to a *technical* egalitarianism, i.e. one that is formulated in terms of von Neumann-Morgenstern utility.

8 Acknowledgements

For useful discussions and/or comments on a previous draft of this paper, I am grateful to Constanze Binder, John Broome, William MacAskill, Toby Ord, and audiences in seminars at Yale, Bristol, the London School of Economics and Erasmus University (Erasmus Institute for Philosophy and Economics).

¹³Parfit writes ‘expectable’, but makes clear that he means what others mean by ‘expected’.

A Extending Technical Prioritarianism to an overall value function on prospects

Section 2 concluded with the observation that the Technical Prioritarian’s defining value function (4) evaluates only states of affairs, and that there is more than one way to extend (4) to a value function on prospects (representing the *ex ante* betterness ordering of those prospects). The consistency constraint is that, since states of affairs are degenerate cases of prospects, the extended value function must reduce to one representing the same ordering of states of affairs as (4) when all states of nature but one receive zero probability. The possibilities include:

$$V^{EPP}(\cdot) = \sum_{ij} p_j f(u_i^{VNM}(j \wedge \cdot)); \quad (10)$$

$$V^{EAP} = \sum_i f \left(\sum_j p_j u_i^{VNM}(j \wedge \cdot) \right); \quad (11)$$

$$V^{EEDEP} = \sum_j p_j \left(\frac{1}{n} \sum_i f(u_i^{VNM}(j \wedge \cdot)) \right). \quad (12)$$

V^{EPP} represents the standard Ex Post Prioritarianism defended by e.g. Parfit (2012)¹⁴ and Porter (2012), according to which von Neumann-Morgenstern utility has diminishing marginal moral value. V^{EAP} represents Ex Ante Prioritarianism, giving priority in improvements in *ex ante expected* von Neumann-Morgenstern utility to the *ex ante* worse off.¹⁵ V^{EEDEP} (‘Expected Equally Distributed Equivalent Prioritarianism’, henceforth EEDE Prioritarianism) is motivated by its avoidance of some of the arguably undesirable features of

¹⁴In the paper cited, Parfit writes that ‘we ought ... to accept’ what he calls ‘the equal chances view’, i.e. the view that ‘when we could save the life of only one of two people, who do not differ in relevant ways, we ought to give these people equal chances of being the person whose life we save’ (p. 431), and that ‘prioritarians should accept’ the view that ‘in deciding what we ought to do, we ought to take into account not only the goodness of outcomes, but also the goodness of people’s prospects, in the sense of their chances of receiving benefits or burdens, and their expectable levels of well-being’ (p. 432). On a superficial reading, these views seems more in keeping with Ex Ante than with Ex Post Prioritarianism. However, the context suggests that Parfit intends to limit the relevance of individuals’ *status quo ex ante* expected goodness to the deontic, rather than the telic, part of his overall moral theory. For instance, one page later, in discussing his ‘Expanded Nine’ case, Parfit writes: “*When we assess the goodness of these outcomes*, it makes no difference whether it is Jack or Jill who might have these longer or shorter lives, so these three acts would have outcomes that would be expectably equally good. But these acts differ in their effects on people’s prospects”; he concludes from this that, according to a principle (his PP3) that prioritarians ‘may believe’, ‘if other things are equal, *we ought to do*’ one of the three acts under discussion rather than either of the other two (my emphasis).

¹⁵‘Ex ante’ views in general, if not this specific form, have been defended by ((Diamond, 1967), (Epstein & Segal, 1992)).

V^{EPP} and V^{EAP} , and is suggested by e.g. Marc Fleurbaey (2010). We pause to note how each of these, in turn, avoids the conditions of Harsanyi’s theorem.

A.1 Ex Post prioritarianism

The Ex Post Prioritarian’s ordering of prospects manifestly satisfies the axioms of decision theory; it violates the Ex Ante Pareto principle. In fact, V^{EPP} violates Ex Ante Pareto even in cases involving only one person.¹⁶ To see the latter, consider, e.g., the following pair of prospects. For concreteness, suppose again that the prioritarian’s transform f is the square root transform: then the ‘ u^{VNM} ’ numbers in the first row give the ‘ $f(u^{VNM})$ ’ numbers in the second row.

Example 4: Ex Post Prioritarianism violates the Ex Ante Pareto principle even in one-person cases.

	Prospect G		Prospect H	
	H	T	H	T
u^{VNM}	0	9	4	4
$f(u^{VNM})$	0	3	2	2

It is the definition of von Neumann-Morgenstern utility u^{VNM} that betterness-for-the-individual tracks expected von Neumann-Morgenstern utility; thus, in this example, prospect G is *ex ante* better for the only person involved than prospect H. However, according to the Ex Post Prioritarian, H is overall better than G, since according to that theory overall betterness tracks, not the expectation value of von Neumann-Morgenstern utility itself, but the expectation value of the square root of von Neumann-Morgenstern utility.

The fact that only one person figures in Example 4 is inessential to the basic point it makes. The following two-person example is if anything even more bizarre:¹⁷

Example 5: Another odd implication of Ex Post Prioritarianism.

	Prospect I		Prospect J	
	H	T	H	T
u_1^{VNM}	36	36	4	100
u_2^{VNM}	49	49	4	100

In this example, not only is J *ex ante* better for both persons (1 and 2): J also has perfect *ex ante* equality, whereas I does not, and J has a guarantee of *ex post* equality, whereas I has a guarantee of *ex post* inequality. Yet (ex post Technical) prioritarrians hold that I is overall better than J.

¹⁶This point was made in (McCarthy, 2008).

¹⁷I am grateful to Toby Ord for this example.

It is thus abundantly clear that the manner in which Ex Post Prioritarianism violates the Ex Ante Pareto principle cannot be rationalised by appeal to any intrinsic value of interpersonal equality.

A.2 Ex Ante Prioritarianism

Of the three value functions discussed here, the *Ex Ante* Prioritarian value function V^{EAP} is the only one that respects the Ex Ante Pareto principle. Instead, it violates one of Harsanyi’s other assumptions: that the betterness-*simpliciter* ordering of prospects conform to standard decision theory. (That is, there is no expected-value representation of the better-than ordering of prospects that is given by V^{EAP} .) Specifically, it violates the Sure Thing principle, which requires that if each person’s *ex post* well-being if the coin lands Heads is held fixed, the ordering of prospects that have varying outcomes under Tails is independent of the precise *levels* at which *ex post* well-being under Heads is held fixed. For instance:

Example 6: Ex Ante Prioritarianism violates the Sure Thing principle.

	Prospect K		Prospect L		Prospect M		Prospect N	
	H	T	H	T	H	T	H	T
u_1^{VNM}	0	0	0	100	100	0	100	100
u_2^{VNM}	100	100	100	0	0	100	0	0

According to Ex Ante Prioritarianism, $L \succ K$, but $M \succ N$. This is a violation of the Sure Thing principle.

Ex Ante Prioritarianism also violates a principle of avoidance of foreseeable regret: it can happen that the theory judges one prospect to be better (*ex ante*) than another, even when it is known that *ex post*, the same theory would rank the two prospects in the opposite order in every state of nature (Fleurbaey & Voorhoeve, n.d.)¹⁸. For example:

Example 7: Ex Ante Prioritarianism violates a principle of avoidance of foreseeable regret.

	Prospect O		Prospect P	
	H	T	H	T
u_1^{VNM}	36	36	25	48
u_2^{VNM}	36	36	48	25

Ex Ante Prioritarianism judges *ex ante* that $P \succ O$, but *ex post* the same theory judges that $O \succ P$ whether the coin lands Heads or Tails.

¹⁸The principle in question is Fleurbaey and Voorhoeve’s “Principle of full information, Part I”.

A.3 EEDE prioritarianism

We turn now to the value function V^{EEDEP} , ‘Expected Equally Distributed Equivalent Prioritarianism’. This is sometimes proposed (e.g. (Fleurbaey, 2010)) as a way of capturing prioritarian-type intuitions while avoiding many of the odd features of V^{EPP} and V^{EAP} noted above. It does reduce to the defining prioritarian form (4) in the absence of risk — thus, it is a species of Technical Prioritarianism by my definition — but in other respects is perhaps more egalitarian than prioritarian (note that the full expression V^{EEDEP} does not exhibit separability of persons; cf. footnote 6). As is easy to verify by inspecting the expression for V^{EEDEP} , this value function is guaranteed to respect the Ex Ante Pareto principle in cases of perfect equality, and *a fortiori* in cases involving only one person; it is only in cases of interpersonal inequality that it may violate Ex Ante Pareto. Thus it *is*, at least in principle, open to an advocate of V^{EEDEP} to defend these violations of Ex Ante Pareto by appeal to the egalitarian rationale outlined above.

For these reasons, V^{EEDEP} strikes me as the most plausible way of extending the Technical Prioritarian value function (4) on riskless prospects to cases of risk, and in particular, as far better than V^{EPP} or V^{EAP} . It is not, however, without its own problems. For instance, as pointed out by Fleurbaey (*ibid.*), it does violate an apparently plausible principle of ‘irrelevance of the utilities of the dead’.

B Extending Technical Cautionism to an overall value function on prospects

At the close of section 3, we had the cautionist’s individual value function (9), together with the overall value function *for states of affairs*, (8). As in the prioritarian case, to complete the Technical Cautionist theory, the remaining task is to extend (8) to an overall value function *on prospects*. The consistency constraint is that this value function for prospects must agree with (8) on the ordering of riskless prospects. The possible extensions of (8) include:

$$\text{Cautionism}_1: V^{c1}(\cdot) = \sum_{ij} p_j u_i^D(j \wedge \cdot);$$

$$\text{Cautionism}_2: V^{c2}(\cdot) = \sum_j p_j g\left(\frac{1}{n} \sum_i u_i^D(j \wedge \cdot)\right);$$

$$\text{Cautionism}_3: V^{c3}(\cdot) = \sum_i g^{-1}\left(\sum_j p_j g\left(u_i^D(j \wedge \cdot)\right)\right).$$

All three versions of cautionism include risk aversion with respect to Debreu utility in their representation (9) of betterness-for-*i*. Cautionism₁, for some reason or none, declines (however) to include any risk aversion with respect to Debreu utility in its representation of betterness-*simpliciter*. Cautionism₂ and Cautionism₃ both include some such risk aversion, but differ on the details. According to Cautionism₂, a prospect is improved by transferring a given number

of Debreu utiles from one state of affairs with a higher total to an equiprobable state of affairs with a lower total. The axiology represented by Cautionism₃ is harder to characterise verbally; like EEDE Prioritarianism, it is motivated (if at all) by the fact that it reduces to the required formula (here, formula (8)) in the absence of risk, and that it lacks some of the arguably undesirable features of other extensions of (8).

Each of the value functions V^{c1}, V^{c2}, V^{c3} represents an axiology that is in genuine conflict with the utilitarian one. Correspondingly, each violates one or more of the axioms of Harsanyi's aggregation theorem. Cautionism₁ and Cautionism₂ violate the Ex Ante Pareto principle; the overall value function postulated by Cautionism₃ violates expected utility theory. These claims are illustrated by the following examples. (For concreteness, I suppose throughout that the cautionist's transform g is the square root transform, as I assumed above for the prioritarian's transform f .)

Example 8: Cautionism₁ violates the Ex Ante Pareto principle.

	Prospect Q		Prospect R	
	H	T	H	T
u^D	0	9	4	4
$g(u^D)$	0	3	2	2

Any cautionist will hold that $R \succ_i Q$ (by (9)). Thus, consistency with the Ex Ante Pareto principle would require that $R \succ Q$. According to Cautionism₁, however, $Q \succ R$.

Example 9: Cautionism₂ violates the Ex Ante Pareto principle.

	Prospect S		Prospect T	
	H	T	H	T
u_1^D	36	0	16	9
u_2^D	0	36	9	16
$g(u_1^D)$	6	0	4	3
$g(u_2^D)$	0	6	3	4

Any cautionist will hold (by 9) that $T \succ_i S$ for $i = 1, 2$. Thus, consistency with Ex Ante Pareto would require that $T \succ S$. But according to Cautionism₂, $S \succ T$.

Example 10: Cautionism₃ violates the Sure Thing principle.

	Prospect U		Prospect V		Prospect W		Prospect X	
	H	T	H	T	H	T	H	T
u_1^D	0	0	0	100	100	0	100	100
u_2^D	100	100	100	0	0	100	0	0

The Sure Thing principle requires that $U \succ V$ iff $W > X$. But according to Cautionism₃, $U \succ V$ but $X \succ W$.

C Value functions

We collect here the various doctrines regarding the relationship between goodness-for- i and goodness-*simpliciter* that have been discussed in Appendices A and B, as represented by their value functions. For each doctrine, we specify (i) the value function V_i representing betterness-for- i , (ii) the value function V representing betterness-*simpliciter* among multi-person lotteries, and (iii) the restriction $V|_{cert}$ of V to cases in which no uncertainty is present. In each case, we do so both in terms of von Neumann-Morgenstern utility (u^{VNM}) and in terms of Debreu utility (u^D). Recall that the expressions for V_i in terms of u^{VNM} and for $V|_{cert}$ in terms of u^D have the status of definitions (of u^{VNM} and u^D respectively), and are therefore agreed on by all theories. It is definitive of Technical Prioritarianism (resp. of Technical Cautionism) that $u^D = f(u^{VNM})$ (resp. that $u^{VNM} = g(u^D)$), where f and g are (unspecified) strictly increasing and strictly concave functions. Combining the last two facts, it follows that all versions of Technical Prioritarianism agree with one another, and all versions of Technical Cautionism agree with one another, over the expression for V_i in terms of u^D and the expression for $V|_{cert}$ in terms of u^{VNM} ; thus, given that a particular doctrine is prioritarian (or that it is cautionist), the only open question is how the theory extends $V|_{cert}$ to risky prospects. *Ex Post* Prioritarianism is the ‘opposite’ of Cautionism₁ (in the sense of being obtainable therefrom via the replacement $f \rightarrow g^{-1}$); *Ex Ante* Prioritarianism of Cautionism₃; EEDE Prioritarianism of Cautionism₂.

Doctrine	Relationship between u^{VNM} and u^D	Value functions, in terms of u^{VNM}	Value functions, in terms of u^D
Utilitarianism	$u^{VNM} = u^D$	$V(\cdot) = \sum_{ij} p_j u_i^{VNM}(j \wedge \cdot)$ $V _{cert} = \sum_i u_i^{VNM}$ $V_i(\cdot) = \sum_j p_j u_i^{VNM}(j \wedge \cdot)$	$V(\cdot) = \sum_{ij} p_j u_i^D(j \wedge \cdot)$ $V _{cert} = \sum_i u_i^D$ $V_i(\cdot) = \sum_j p_j u_i^D(j \wedge \cdot)$
Ex Post prioritarianism	$f(u^{VNM}) = u^D$	$V(\cdot) = \sum_{ij} p_j f(u_i^{VNM}(j \wedge \cdot))$ $V _{cert} = \sum_i f \circ u_i^{VNM}$ $V_i(\cdot) = \sum_j p_j u_i^{VNM}(j \wedge \cdot)$	$V(\cdot) = \sum_{ij} p_j u_i^D(j \wedge \cdot)$ $V _{cert} = \sum_i u_i^D$ $V_i(\cdot) = \sum_j p_j f^{-1}(u_i^D(j \wedge \cdot))$
Ex Ante prioritarianism	$f(u^{VNM}) = u^D$	$V(\cdot) = \sum_i f\left(\sum_j p_j u_i^{VNM}(j \wedge \cdot)\right)$ $V _{cert} = \sum_i f \circ u_i^{VNM}$ $V_i(\cdot) = \sum_j p_j u_i^{VNM}(j \wedge \cdot)$	$V(\cdot) = \sum_i f\left(\sum_j p_j f^{-1}(u_i^D(j \wedge \cdot))\right)$ $V _{cert} = \sum_i u_i^D$ $V_i(\cdot) = \sum_j p_j f^{-1}(u_i^D(j \wedge \cdot))$
EEDE Prioritarianism	$f(u^{VNM}) = u^D$	$V(\cdot) = \sum_j p_j f^{-1}\left(\frac{1}{n} \sum_i f(u_i^{VNM}(j \wedge \cdot))\right)$ $V _{cert} = \sum_i f \circ u_i^{VNM}$ $V_i(\cdot) = \sum_j p_j u_i^{VNM}(j \wedge \cdot)$	$V(\cdot) = \sum_j p_j f^{-1}\left(\frac{1}{n} \sum_i u_i^D(j \wedge \cdot)\right)$ $V _{cert} = \sum_i u_i^D$ $V_i(\cdot) = \sum_j p_j f^{-1}(u_i^D(j \wedge \cdot))$
Cautionism ₁	$u^{VNM} = g(u^D)$	$V(\cdot) = \sum_{ij} p_j g^{-1}(u_i^{VNM}(j \wedge \cdot))$ $V _{cert} = \sum_i g^{-1} \circ u_i^{VNM}$ $V_i(\cdot) = \sum_j p_j u_i^{VNM}(j \wedge \cdot)$	$V(\cdot) = \sum_{ij} p_j u_i^D(j \wedge \cdot)$ $V _{cert} = \sum_i u_i^D$ $V_i(\cdot) = \sum_j p_j g(u_i^D(j \wedge \cdot))$
Cautionism ₂	$u^{VNM} = g(u^D)$	$V(\cdot) = \sum_j p_j g\left(\frac{1}{n} \sum_i g^{-1} u_i^{VNM}(j \wedge \cdot)\right)$ $V _{cert} = \sum_i g^{-1} \circ u_i^{VNM}$ $V_i(\cdot) = \sum_j p_j u_i^{VNM}(j \wedge \cdot)$	$V(\cdot) = \sum_j p_j g\left(\frac{1}{n} \sum_i u_i^D(j \wedge \cdot)\right)$ $V _{cert} = \sum_i u_i^D$ $V_i(\cdot) = \sum_j p_j g(u_i^D(j \wedge \cdot))$
Cautionism ₃	$u^{VNM} = g(u^D)$	$V(\cdot) = \sum_i g^{-1}\left(\sum_j p_j u_i^{VNM}(j \wedge \cdot)\right)$ $V _{cert} = \sum_i g^{-1} \circ u_i^{VNM}$ $V_i(\cdot) = \sum_j p_j u_i^{VNM}(j \wedge \cdot)$	$V(\cdot) = \sum_i g^{-1}\left(\sum_j p_j g(u_i^D(j \wedge \cdot))\right)$ $V _{cert} = \sum_i u_i^D$ $V_i(\cdot) = \sum_j p_j g(u_i^D(j \wedge \cdot))$

References

- Adler, M., & Sanichirico, C. (2006). Inequality and uncertainty: Theory and legal applications. *University of Pennsylvania Law Review*, 155, 279–377.
- Broome, J. (1991). *Weighing goods*. Oxford: Blackwell.

- Debreu, G. (1960). Topological methods in cardinal utility theory. In S. K. Kenneth J. Arrow & P. Suppes (Eds.), *Mathematical methods in the social sciences*. Stanford University Press.
- Diamond, P. A. (1967, October). Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility: Comment. *Journal of Political Economy*, 75(5), 765–6.
- Epstein, L. G., & Segal, U. (1992, August). Quadratic social welfare functions. *Journal of Political Economy*, 100(4), 691–712.
- Fleurbaey, M. (2010, August). Assessing risky social situations. *Journal of Political Economy*, 118, 649–680.
- Fleurbaey, M., & Voorhoeve, A. (n.d.). Decide as you would with full information! An argument against ex ante Pareto. In N. Eyal, S. Hurst, O. Norheim, & D. Wikler (Eds.), *Health inequality: Ethics and measurement*. Oxford University Press. (Forthcoming)
- Greene, J., & Baron, J. (2001). Intuitions about declining marginal utility. *Journal of behavioural decision making*, 14, 243–55.
- Harsanyi, J. (1955, August). Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy*, 63(4), 309–321.
- McCarthy, D. (2008, March). Utilitarianism and prioritarianism ii. *Economics and Philosophy*, 24, 1–33.
- Nagel, T. (1979). Equality. In *Mortal questions*. Cambridge University Press.
- Nagel, T. (1986). *The view from nowhere*. Oxford University Press.
- Otsuka, M., & Voorhoeve, A. (2009). Why it matters that some are worse off than others: An argument against the priority view. *Philosophy and Public Affairs*, 37, 171–99.
- Parfit, D. (1995). *Equality or priority?* Department of Philosophy, University of Kansas. (The Lindley Lecture, University of Kansas, November 21, 1991.)
- Parfit, D. (1997, December). Equality and priority. *Ratio*, 10, 202–222.
- Parfit, D. (2012, September). Another defence of the priority view. *Utilitas*, 24, 399–440.
- Porter, T. (2012, September). In defence of the priority view. *Utilitas*, 24, 349–364.
- Rabinowicz, W. (2002). Prioritarianism for prospects. *Utilitas*, 14(1), 2–21.
- Rawls, J. (1972). *A theory of justice*. Oxford University Press.
- Scanlon, T. (1975). 'preference and urgency. *The Journal of Philosophy*, 72(19), 655–669.
- Scheffler, S. (1982). *The rejection of consequentialism*. Oxford: Clarendon Press. (Revised edition 1994)
- Sen, A. (1973). *On economic inequality*. Oxford University Press.
- Sen, A. (1976). Welfare inequalities and rawlsian axiomatics. *Theory and decision*, 7, 243–62.
- von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behaviour*. Princeton University Press.

- Williams, A. (2012, September). The priority view bites the dust? *Utilitas*, 24, 315–331.
- Williams, B. (1973). A critique of utilitarianism. In J. J. C. Smart & B. Williams (Eds.), *Utilitarianism: For and against*. Cambridge University Press.