A New Counterexample to Prioritarianism

TOBY ORD
University of Oxford

Prioritarianism is the moral view that a fixed improvement in someone's well-being matters more the worse off they are. Its supporters argue that it best captures our intuitions about unequal distributions of well-being. I show that prioritarianism sometimes recommends acts that will make things more unequal while simultaneously lowering the total well-being and making things worse for everyone ex ante. Intuitively, there is little to recommend such acts and I take this to be a serious counterexample for prioritarianism.

In 'Equality and Priority', Derek Parfit introduced prioritarianism as a way of valuing distributions of well-being across individuals according to which 'benefiting people matters more the worse off these people are'.[1] Unlike utilitarianism, prioritarianism is sensitive to the distribution of well-being in a population. For the same total well-being, prioritarianism judges an outcome to be better when the well-being is more equally distributed. Its advocates consider it to be a particularly principled way of reaching such a conclusion as it does not need to say that it is better in any respect if we make a distribution more equal simply by levelling down everyone's well-being. While the name suggests a form of consequentialism in which well-being is the only thing that matters morally, everything I shall say would also apply to theories in which well-being is just one part of morality, so long as that part should be evaluated with priority to the worse-off.

There are many different varieties of prioritarianism. One key question is exactly how much more important it is to benefit someone who is worse off. There are infinitely many different ways to answer this, with each way corresponding to a different mathematical function. Let us use the term w_1 to denote the well-being of person p_1 and so forth. Utilitarianism says that the value of the entire population is just the

[1] D. Parfit, 'Equality and Priority', *Ratio*, NS 10 (1997), pp. 202–21. This article arose from Parfit's Lindley Lecture of 1991 (published as D. Parfit, 'Equality or Priority?', *The Ideal of Equality*, ed. M. Clayton and A. Williams (Basingstoke, 2000), pp. 347–86. Versions of this idea have been examined much earlier within the study of social welfare functions in economics. See e.g. A. Atkinson and J. E. Stiglitz, *Lectures on Public Economics* (London, 1980), p. 340.

© Cambridge University Press 2015
Utilitas Vol. 27, No. 3, September 2015
doi:10.1017/S0953820815000059

CrossMark
https://doi.org/10.1017/S0953820815000059 Published online by Cambridge University Press

sum of these well-being levels. For a population of n people we have:

$$\text{Utilitarianism}: \quad \text{value} = w_1 + w_2 + w_3 + \ldots + w_n$$

In contrast, prioritarianism applies a weighting function to each person's well-being before summing them. This function is chosen to be strictly increasing (so that higher well-beings are better), but to have a slope that is becoming ever flatter (so that the moral value of improvements to someone's well-being matter less the better off the person is). This is known as a concave function. Commonly used examples include square root and logarithm. More generally, we can simply denote the function as f.

$$\text{Prioritarianism}: \quad \text{value} = f(w_1) + f(w_2) + f(w_3) + \ldots + f(w_n)$$

Not all forms of inequality-averse aggregation can be expressed by the above formula. This is because some forms have additional interdependence between different people's well-being levels. For instance the choice of how best to allocate well-being between persons 1 and 2 might depend on person 3's level of well-being. Such theories are said to be non-separable.[2] Following John Broome,[3] I take the term 'prioritarianism' to refer only to the separable theories and to use the term 'egalitarianism' to refer to the non-separable ones.

Another key question concerns how to evaluate uncertain prospects. Two different approaches have emerged.[4] One is to determine first the prioritarian value of each outcome that could occur using the formula above and then take the expected value of these. This is known as *ex post prioritarianism* and it coheres best with standard approaches to decision theory. The other is *ex ante prioritarianism* in which we first determine the expected well-being of each individual and then use these numbers in the formula above. The conclusions of this article only apply to ex post prioritarianism. Ex ante prioritarianism suffers from its own serious problems, but they are not the topic of this article.[5]

Finally, Parfit's original lecture allowed for an interpretation of prioritarianism in which it only applied in interpersonal cases and

[2] For formal definitions of several different forms of separability, see J. Broome, *Weighing Goods* (Oxford, 1991), pp. 60–89. What I refer to here is also called *strong separability* and is provably equivalent to *additive separability*.

[3] J. Broome, 'Equality versus Priority: A Useful Distinction', *Fairness and Goodness in Health*, ed. D. Wikler and C. Murray (Geneva, 2003).

[4] See D. McCarthy, 'Utilitarianism and Prioritarianism II', *Economics and Philosophy* 24 (2008), pp. 1–33.

[5] See M. D. Adler, *Well-being and Fair Distribution* (New York, 2012), ch. 7.

https://doi.org/10.1017/S0953820815000059 Published online by Cambridge University Press

not in intrapersonal cases.[6] The results of this article do not apply to such an interpretation, but are levelled at the most widely held version, where the same method of aggregating benefits is applied regardless of how many people's interests are at stake.

We now have enough detail to look at some potential counterexamples to the view. Rabinowicz has presented a very challenging case in which prioritarianism prefers the act which is worse for everyone from an ex ante perspective.[7] That is, it has a lower expected well-being for all individuals, and yet it is preferred by prioritarianism. It might be acceptable for a theory to prefer an act with a lower expected well-being for all individuals if it also held some striking advantage — for example, deontologists would typically accept this if the only alternative involved violating a deontological constraint. However, Rabinowicz has shown how prioritarianism can recommend lowering everyone's expected well-being even in the absence of such deontic considerations. I find this to be a very implausible implication for a moral theory to have, as explained in some detail by David McCarthy.[8]

The only possibly redeeming feature I can see in Rabinowicz's example is that the act preferred by prioritarianism is guaranteed to produce a more equal outcome ex post. While thoroughgoing prioritarians are not supposed to care about equality per se, this still affects the intuitive force of the example. Many people who are now prioritarians preferred more equal distributions of well-being before becoming familiar with the non-relational argument for doing so, and saw prioritarianism as a way of explaining that moral preference while avoiding giving any support to equality through levelling down. Several prioritarians whom I have personally asked about Rabinowicz's example have appealed in some manner to the increased equality in the act preferred by prioritarianism, for example by framing it as just another case of an equality–efficiency trade-off. Indeed it is difficult to rule out some round-about moral explanation through which the increased equality could compensate for the loss of expected well-being to everyone, while keeping close to at least some of the core values of prioritarianism.

However, building on Rabinowicz's example, I present a new and stronger counterexample which closes any hope of such an escape. For the act preferred by prioritarianism is both worse for everyone ex ante and *more* unequal, making it intuitively extremely unappealing. The

[6] See Parfit, 'Equality and Priority', p. 213. For further discussion of such interpretations of prioritarianism, see A. Williams, 'The Priority View Bites the Dust?', *Utilitas* 24 (2012), pp. 315–31.

[7] W. Rabinowicz, 'Prioritarianism for Prospects', *Utilitas* 14 (2002), pp. 2–21.

[8] McCarthy, 'Utilitarianism and Prioritarianism II'.

https://doi.org/10.1017/S0953820815000059 Published online by Cambridge University Press

Table 1

	Heads		Tails		Expectation		Prioritarian
	p_1	p_2	p_1	p_2	p_1	p_2	Valuation
Act A	36	49	36	49	36	49	13
Act B	4	4	100	100	52	52	12

counterexample involves two available acts (A and B) which affect two people (p_1 and p_2). Act A has a certain outcome (36 units of well-being for p_1 and 49 units for p_2). Act B will either lead to 4 units of well-being for both people or 100 units for both people, with equal chances of each. For simplicity I consider this to be the result of a coin flip and call the states that could arise *Heads* and *Tails*. This is represented by the left-hand side of table 1 (up to the double line).

In the right hand part of the table I have summarized the expected well-being of each person under each act and also the overall moral value that prioritarianism assigns to each act. For clarity I have set up the example to work with a particular prioritarian weighting function. In this case it is square root. Versions of this example can be set up for any prioritarian weighting function.[9]

As we can see, act A is ranked as superior by prioritarianism. However, it has a number of very unfavourable properties. It has less total expected well-being (85 vs 104). It is also worse for everyone ex ante. Moreover, while Rabinowicz's example arguably compensated for these defects with the prioritarian act leading to a more equal outcome ex post, here it is guaranteed to increase inequality in terms of both ex ante and ex post well-being.

Why is prioritarianism recommending such an apparently unattractive act? It is because the prioritarian weighting function f makes prioritarianism risk averse about well-being as well as inequality averse. As well as caring more about a given benefit if it goes to a badly off person rather than a well off person, it cares more about a given benefit if it goes to a person in a state of nature where they are badly off, rather than one in which they are well off (in this case the Heads state). Indeed the weighting function makes prioritarianism

[9] For example, if the function were $\log_2$ we could replace the numbers 4, 36, 49, 100 with 1, 4, 8, 16. In general we have four numbers which we shall call a, b, c, d. For the example to work, we require that $a < b < c < d$, $b + c < a + d$, and $f(b) + f(c) > f(a) + f(d)$. This can be achieved for any f if we choose a to be less than d, then set b to be 2/3 a + 1/3 d − ε and set c to be 1/3 a + 2/3 d − ε, where ε is a small number that has to be closer to zero the closer f is to linear.

https://doi.org/10.1017/S0953820815000059 Published online by Cambridge University Press

averse to risk and inequality to exactly the same degree (set by the choice of f).

Prioritarians may not have been too daunted by endorsing a view that is risk averse about well-being, but this example shows that this commitment to risk aversion has serious bite. The counterexample was constructed to make act A worse in all ways except for satisfying risk aversion. Prioritarianism is recommending it on grounds that do fit with its foundational principle – if benefits matter more the worse off you are, then they matter more in states of nature when you are badly off – but which come at a high cost to the intuitive principles of making people better off and avoiding unequal interpersonal distributions.

What this example does is to bring prioritarianism's inbuilt risk aversion to the fore, showing that it can compete with its inequality aversion. If there is enough risk at stake, then in order to avoid it prioritarianism can recommend an act that:

- has less total well-being ex ante
- is worse for everyone ex ante
- is more unequal ex ante
- is more unequal ex post

Rabinowicz's counterexample could be seen as showing that prioritarianism endorses a form of ex ante levelling down – lowering everyone's expected well-being while increasing equality. This new counterexample involves moving individuals' expected well-being down in a way that also makes things more unequal. This makes it clear that prioritarianism does not merely recommend trade-offs between efficiency and equality. In some cases it recommends a reduction in both. I take it that this result would be surprising to most (ex post) prioritarians and a conclusion that they would be reluctant to endorse.[10]

toby.ord@philosophy.ox.ac.uk

[10] I wish to acknowledge the Oxford Martin School for funding this research and Marc Fleurbaey, Andrew Lister, Derek Parfit, Larry Temkin, Alex Voorhoeve and an anonymous referee for helpful discussion.

https://doi.org/10.1017/S0953820815000059 Published online by Cambridge University Press