

Competing interests: The author(s) declare none

This is a pre-corrections draft; please cite the final published paper.

The Prospects for ‘Prospect Utilitarianism’

Ben Davies, Uehiro Centre for Practical Ethics, University of Oxford

benjamin.davies@philosophy.ox.ac.uk

Hun Chung argues for a theory of distributive justice—‘prospect utilitarianism’—that overcomes two central problems purportedly faced by sufficientarianism: giving implausible answers in ‘lifeboat cases’, where we can save the lives of some but not all of a group, and failing to respect the axiom of continuity. Chung’s claims that prospect utilitarianism overcomes these problems, and receives empirical support from work in economics on prospect theory. This paper responds to Chung’s criticisms of sufficientarianism, showing that they are misplaced. It then shows that prospect utilitarianism faces independent problems, since it too requires a threshold, which Chung bases on the idea of ‘adequate functioning’. The paper shows that there are problems with this as a threshold, and that it is not empirically supported by prospect theory.

Introduction

Hun Chung (2017) has recently developed an interesting view of distributive justice, ‘prospect utilitarianism’ (PU). In Chung’s view, PU can accommodate the central intuitive insights of sufficientarianism without its weaknesses. Chung identifies two central problems for sufficientarianism: it gives counter-intuitive answers to ‘lifeboat’ cases (pp.1913-1915), and it is not ‘continuous’, since it allows small differences in the welfare levels of a population to make big differences in our ethical assessment of that bringing about that population (pp.1915-1916). Chung elicits empirical support (pp.1920-1921) for PU from behavioural economics. Kahneman and Tversky (1979) conducted various experiments that suggest people’s preferences about losses and gains are governed by ‘prospect theory’. Of central relevance is their finding that “the aggravation that one experiences in losing a sum of money appears to be greater than the pleasure associated with gaining the same amount”. They summarise their findings thus:

we have proposed that the [utility function] is (i) defined on deviations from the reference point; (ii) generally concave for gains and commonly convex for losses; (iii) steeper for losses than for gains.

(Kahneman and Tversky 1979: 279)

In other words:

- people respond differently to gains and losses, relative to some reference point.
- It is better to get a given sum of money when close to the threshold than when far away, moving both up and down from the reference point.

- The change in utility from a change of resource of size n is greater when below the reference point than when above.

Chung's PU is a utilitarian view which adopts the observations of prospect theory. Each individual has an "individual critical sufficiency threshold", which differs from others' thresholds, namely "the amount of material resources he/she needs to adequately function as a normal human being" (p.1921). Their personal utility function obeys prospect theory with respect to that threshold: losses are more negative than equivalent gains are positive; and additional resources are worth more the closer the individual is to the threshold in either direction. PU is utilitarian in advocating that we maximise total utility. Since individuals comply with prospect theory, social utility maximisation will have egalitarian distributive effects: resources do more good the closer the individual is to adequate functioning, and do more good ensuring that people do not fall below this threshold than in securing equivalent advances above it.

This paper outlines how Chung's criticisms of sufficientarianism are misplaced. It also argues that PU faces theoretical problems of its own, and that it receives little empirical support from prospect theory.

Lifeboat cases

Sufficientarianism's best-known articulation is from Frankfurt (1987), who argues that sufficiency is preferable to equality when we face two options:

Some Survive: Save the lives of some, but not all, of a group by allocating resources unequally;

All Die: Allocate resources equally, consigning everyone to death.

The better option is Some Survive.

Frankfurt's view is a 'headcount' approach, which maximises the number of individuals are above a sufficiency threshold, but demands nothing for people who are already above this threshold, or for those unavoidably below it. Headcount sufficientarians face a problem with lifeboat cases. If our sufficiency threshold is survival, headcounters favour Some Survive. But since they are unconcerned about those above the threshold, setting the threshold at bare survival says that the only concern of justice is that as many people survive as possible. An alternative is setting the threshold higher. Once everyone has enough resources for a fully flourishing life, it is more plausible that they have no further claims of justice. However, as Chung notes, headcount sufficientarianism with a high threshold tells us to be unconcerned with those who will remain below the threshold no matter what.

Our lifeboat passengers will likely not end up above the threshold of a very comfortable life, and so high-threshold headcount sufficientarianism cannot recommend Some Survive over All Die.

Chung considers two non-headcount sufficientarian views. Crisp (2003) thinks those below the threshold should have absolute priority: we should prioritise any benefit to someone below the threshold over any benefit to someone above it. Among those below the threshold, we should have non-absolute prioritarian concern for the worse off. Huseby (2010) advocates ‘subsistence’ (the point at which one’s basic needs are met) as a threshold. However, he adopts a multiple threshold view, which includes an upper limit of ‘full contentment’. Different principles apply depending on the relevant individuals’ positions relative to each threshold.

Chung thinks both views fail on lifeboat cases, at least because they are insufficiently specific. Since all lifeboat passengers are (at the time of distributive decision) below Huseby’s threshold of subsistence, his view says that they should be given either an absolute or ‘strong’ priority over those who are above those thresholds. However, it does not tell us how to decide *between* individuals below the minimal threshold.

In contrast, says Chung, PU advocates Some Survive in lifeboat cases because it prioritises those who are below their personal threshold compared with those who are above theirs, but among the former category prioritises those who can most easily cross the threshold. In lifeboat situations all are below their critical threshold, so PU recommends bringing as many as possible above the threshold.

Lifeboats and sufficiency

An important development in sufficientarianism has been recognising that a concern with sufficiency does not require absolute priority to the worst off, even among those below the threshold. Giving priority to the worse off below the threshold is consistent with caring about benefit size (Shields 2018; Huseby 2020). Indeed, in the very paper Chung cites, Huseby claims it is bad when people do not have enough, and worse “the farther from a sufficient level a person is...and worse the more people that are not sufficiently well off” (2010: 180). This latter specification provides reason to prefer Some Survive. Similarly, Crisp gives explicit instruction regarding those who are below the threshold: “Below the threshold, benefiting people matters more the worse off those people are, the more of those people there are, and the greater the size of the benefit in question” (2003: 758). Puzzlingly, Chung cites but does not discuss this quotation. He might think the claim that ‘benefitting people matters more the worse off they are’ presents an issue. Chung assumes that if we choose Some Survive, those who are *not* given resources and thus die are the worst off, and hence on sufficientarian views should be given higher priority. Yet,

says Chung, “in order to give the right answer...the distributional principle must give zero priority to the worst-off patient” (ibid. 1914-15).

It is not obvious that we should judge who is worst off by looking at how people fare *post*-distribution (see Herlitz and Eyal Forthcoming) rather than pre-distribution; and pre-distribution, everyone in lifeboat cases is equally badly off, at least if we focus on their immediate situation. Still, even if we use a post-distribution approach, sufficientarianism has the theoretical resources to cope with this.

If giving ‘zero priority’ to those who will die means giving *zero weight* to their claims, this only makes sense if they will not benefit from allocatable resources at all; if they would benefit from the resources (e.g., through a less painful death), their claims have some weight. Sufficientarians care primarily about the allocation of benefits, not resources. Resources are a means to achieving benefits. So, if someone would not benefit from resources, they have no claim to them no matter how badly off they are. On the other hand, if ‘give zero priority’ is just a shorthand for saying we should choose Some Survive, Crisp (and many other sufficientarians) thinks benefit size also matters; thus, it is open to Crisp to say that benefit size in Some Survive outweighs the fact that more individuals below the threshold receive a benefit in All Die.

Chung indirectly acknowledges this possibility (ibid. 1914) but mischaracterises it as a case where sufficientarians (and egalitarians or prioritarrians) allow decisions which counter their general distributive rules in “exceptional cases”, saying that this is *ad hoc* because it would not accommodate lifeboat cases “as a matter of principle”. But this mischaracterises how sufficientarians typically view giving weight to benefit size; the idea of allowing benefit size to outweigh the importance of benefits going to the worst off is not an *ad hoc* exception, but rather a principle that applies in all cases. A sufficientarian justification of Some Survive need not be *ad hoc* or unprincipled. Thus, I conclude that sufficientarians have the theoretical resources to recommend Some Survive over All Die.

Lifeboats and prospect utilitarianism

PU chooses Some Survive because, while it gives priority to those below the sufficiency threshold over those above it, it prioritises the *better* off among people below the threshold. As such, Chung argues, when we set the threshold as “the amount of material resources he/she needs to adequately function as a normal human being” (2017: 1921), we get the right result.

However, a central question in distributive justice is the timeframe over which particular views of justice apply (e.g., McKerlie 2012). It is unclear whether the ‘adequate’ functioning that grounds Chung’s threshold of critical

sufficiency is supposed to include a 'lifespan' component (where individuals need to function only for a certain amount of time, such as eighty years), or whether it is only concerned with what we might call quality of life, i.e., functioning at particular moments in time. Endorsing one does not necessitate endorsing the other; you might think there is a level of adequate functioning with respect to the phenomenal quality of one's life at *particular* times (e.g., you are not in severe pain, and do not have your opportunities and capacities severely debilitated), without accepting an upper age limit, or set of life experiences, that constitutes adequate 'lifetime' functioning.

Assume Chung intends his view to apply only to the present. Imagine a case where we can spend resources on curative medicine for Hamza, or preventive medicine for Ilse. Hamza suffers a temporary condition that means he is currently not functioning adequately, but which will clear up of its own accord in a month. Ilse is perfectly well now, but has a degenerative condition that, in a year's time, will leave her permanently unable to function adequately. If PU focuses only on the present, it tells us to cure Hamza when we should prevent Ilse's illness.

On the other hand, if Chung wants adequate functioning to have a lifetime scope, it may lead to issues regarding lifeboat cases. Assume eighty years is the amount of time required for 'adequate functioning'. We can save the life of either one 60-year-old, or one ten-year-old. Due to a tragic congenital condition, the ten-year-old will only make it to 30, while each 60-year-old will reach eighty. According to a lifetime version of PU, we should abandon the ten-year-old: her twenty years, being far below the 'lifetime' threshold of critical sufficiency, are worth less than the twenty years that each older passenger gets, which brings them right up to the critical threshold. While we might disagree on the details of this case, many would regard this as an unjust result (Bognar 2015). Either way, PU faces intuitive problems too.

A possible response to this issue is to deny that this is what lifetime PU really implies. After all, PU is based on the empirical claim that we *value* gains when we are below a critical threshold more than when we are above it. But, one might protest, our sixty-year-olds will not value their additional twenty years more than the ten-year-old will value hers.

An immediate question is: 'why not?'. People have all sorts of idiosyncratic preferences. Some people discover a new lease of life in their sixties, while some ten-year-olds do not really think about, and hence do not value, the future that much. In any case, this response highlights a problem with Chung's reliance on prospect theory, and its application to lifeboat cases. Assume that Some Survive is the right choice in lifeboat cases. The fact that the alternative is All Die is a central reason that it seems plausible to prioritise those who are easiest to rescue, as PU does. However, Chung's critical threshold is not survival, but the level of resources required for adequate

functioning as a normal human. This raises the possibility of further cases where PU gives less intuitively compelling recommendations than in standard lifeboat cases. For instance, imagine two people, neither of whom can function adequately (since Chung offers no details of what this means, we cannot get much more specific). Without an investment of resources, Josiah will be moderately far from adequate functioning, and Kate will die. If given resources, Josiah will come much closer to adequate functioning, but will not actually reach that threshold. If *she* receives the resources, Kate will survive, and reach the same level of welfare/distance from adequate functioning that Josiah occupies *without* treatment.

According to prospect utilitarianism, we are morally obligated to improve Josiah's situation, despite the fact that Kate is significantly worse off, and will die without treatment. Quite how intuitive this is depends on our understanding of adequate functioning, raising similar issues to the high/low threshold problem for sufficientarians. If adequate functioning is a demanding concept, Josiah might be quite far from his critical threshold, and yet fairly well off. This would make the claim that we must benefit him rather than saving Kate deeply inequalitarian. What's more, it makes it questionable to say that, among those below their thresholds, the better off always get more utility from a set of resources. If adequate functioning is undemanding, the set of people below the critical sufficiency threshold will be much smaller. It will also, I think, make it more plausible to apply 'inverse priority' below the threshold, because anyone who fails to make it over the threshold will either die, or have an extremely bad life. However, as I will later suggest, this weakens the empirical support for the theory.

Continuity and sufficiency

Chung's second objection to sufficientarianism is its failure to respect 'continuity'. Briefly, and informally, continuity between two variables means there are no "jumps" (p.1926), where we have a small change in one variable, but a large change in another. If ethical preferences are continuous with respect to total population welfare, we are almost indifferent between two outcomes that contain almost the same total welfare. Chung, following Roemer, describes this latter claim—continuity between welfare and ethical value—as an *a priori* 'axiom' of distributive justice (p.1915). According to Roemer (2004: 272), continuity implies "that two social states that are almost the same, in terms of the welfare levels of society's members, must be viewed as almost ethically indifferent". Call this 'welfarist continuity'. Sufficientarianism violates welfarist continuity. Two people can be very close in welfare terms, and yet regarded very differently by sufficientarians if one is above, and the other below, the sufficiency threshold.

PU ranks outcomes continuously according to total utility. Insofar as it overlaps with the recommendations of egalitarian, prioritarian or sufficientarian views, this is derivative on an empirical claim that additional resources increase a person's welfare as hypothesised, i.e., with increasing marginal utility below the level threshold of resources required to "adequately function as a normal human being" (p.1921), and with decreasing marginal utility above.

PU also assumes a continuous relationship between resources and utility. Each individual's utility function is assumed to be continuous (a small increase in resources always improves one's welfare, but the rate of change never 'jumps'), and so the sum of all individuals' utility functions ('total utility') is also continuous, such that "very slight changes in the final distribution of resources will result in only a very slight change...in the total sum of individual welfare" (p.1926). Since it is utilitarian, it assumes a continuous relationship between total utility and ethical preferability: the ethically best outcome is that with the greatest total utility, and two states that differ only slightly in total welfare differ only slightly in terms of ethical preferability.

The case against welfarist continuity

Welfarist continuity as expressed in Roemer and Chung's axiom is much narrower than the Aristotelian view that we should treat like cases alike, and dissimilar cases differently "proportionately to their differences" (Goodin 1999: 189). The welfarist axiom is a view about the relationship between *welfare* and ethical value. Hence, rather than continuity *per se*, the axiom concerns what I have called 'welfarist continuity'.

Continuity *per se* is an attractive feature of an ethical view if understood as saying that, where there is only a slight difference *across all ethically relevant features* between two outcomes, this should not make a big difference to our ethical preferences. But we can clearly embrace a principle of continuity understood in this way, while rejecting *welfarist* continuity, since two outcomes may differ only slightly in welfare, but differ more significantly on some other ethically relevant factor. For instance, egalitarians will strongly prefer a distribution A which has identical population welfare to distribution B, but where the welfare is much more equally distributed.

In discussing continuity, Chung goes further than the idea of welfarist continuity proposed by Roemer; he suggests not only that welfare has a continuous relationship with moral value, but also that it has a continuous relationship with resources (ibid. p.1926). This would mean that there is also a continuous relationship between resources and ethical preferability. Yet it is implausible to think that small changes in resources cannot make big differences to welfare, and thus (even for those who accept welfarist continuity) to ethical preferability. The difference between having enough medicine to save someone's life and being just short of that amount is a small *resource* difference.

The difference between having enough money to rent a room for the night and being just short may be tiny. Yet crossing these resource thresholds can make enormous differences to welfare.

Still, this will not help sufficientarians. As Shields (2018: 34-40) argues, sufficientarianism is only an attractive theory if the idea of a threshold plays an *indispensable* part in our best theory of justice, not merely an instrumental route to some more fundamental goal. While sufficientarianism applied to resources might be plausible, it is clearly not irreducible and indispensable, because the attractiveness of applying thresholds to individual resources can be explained by that being the best way to achieve a more fundamental theoretical goal, such as equality or utility maximisation.

However, sufficientarians can plausibly reject welfarist continuity while embracing the less specific principle of continuity I have suggested. For instance, Shields (2018: 44-81) suggests that one candidate for a sufficientarian threshold is autonomy, and that we have especially weighty reasons to secure for people a level of autonomy sufficient to develop and pursue their own view of the good. Having greater powers of autonomy above this threshold is good for people; but the weight of our reasons to secure such increases is weaker, and discontinuously so. If welfarist continuity is an axiom, this seems to rule out by fiat any view that is interested in autonomy except insofar as it impacts welfare. Take, for example, *On Liberty* (1859/2003), where Mill says we should prefer that people have significant autonomy over their own lives, because a life you choose for yourself is better than one chosen for you. Mill's view is welfarist: the betterness of the life is understood in terms of its welfare or prudential value. But we might instead think that self-chosen lives are *non-prudentially* better even if they contain less welfare, simply because they are lives which individuals choose for themselves. On this view, it is possible to have two identical distributions of welfare, one paternalistically imposed and the other autonomously chosen, and prefer the latter for non-welfarist reasons. Clearly, none of this constitutes a decisive argument for such a view. My point is only that to describe welfarist continuity as *axiomatic* seems to rule out the very possibility of this kind of view being correct. That is unjustified.

Alternatively, sufficientarians might draw a distinction between 'ethical attractiveness' and what is demanded by justice. Lasse Nielsen (2019) suggests that there are different sources of value, and that the values that are relevant to justice are *satiabile*; once the relevant value has been sated, people can become better off, but not in a way that is relevant to justice. What justice demands—and sufficientarianism is a theory of *justice*—may be different from the 'moral value' of a total distribution. Thus, even if small differences in welfare necessarily implied only small

differences in the ethical preferability of an outcome (which is Chung's, and Roemer's, informal characterisation of welfarist continuity), such a relationship may not hold between welfare and the demands of justice.

I conclude that welfarist continuity is not axiomatic. As such, prospect utilitarianism's necessary compliance with welfarist continuity is not the advantage that Chung supposes. That concludes my defence of sufficientarianism against the twin charges that it is non-continuous, and that it fails to properly respond to lifeboat cases.

Setting the threshold: the empirical evidence for prospect theory

Prospect utilitarianism shares some positive elements of sufficientarianism (people who do not have enough are prioritised over those who do) and prioritarianism (when everyone is above their critical sufficiency threshold, benefits are more valuable for those who are closer to their threshold). But it does so only derivatively, on the assumption that individual preferences are such that following these principles will maximise total utility. The empirical basis for that claim is prospect theory.

However, prospect theory does not support PU. Remember that PU offers intuitive results (notwithstanding the issues raised above about PU's implications for other types of cases) because the threshold is set at 'adequate functioning'. So, to retain both intuitive plausibility and its evidential basis, it should be the case that prospect theory tells us that people's preferences 'flip' with respect to the value of benefits when they reach the point of being able to function adequately *and* at no other point. For instance, if there are some people whose utility curve becomes shallower (i.e., further resource gains offer weaker welfare benefits) not when they reach adequate functioning, but when they reach luxury, we will find that PU tells us to prioritise *those* individuals over the worse off, because they are below their threshold of critical sufficiency. This is in some ways equivalent to the problem of 'expensive tastes' that troubled egalitarian theories which take welfare as their currency (Dworkin 1981). Similarly, it would be a serious problem for PU's egalitarian credentials if people's preferences were *adaptive*, such that their threshold shifted when they had adjusted to a new lifestyle, or even when made to believe that they deserved or needed a better lifestyle, e.g., through advertising.

Unfortunately for prospect utilitarianism, this is exactly what prospect theory tells us. For instance, in the very paper that Chung cites, we find Kahneman and Tversky telling us that "the same level of wealth...may imply abject poverty for one person and great riches for another – depending on their current assets" (1979: 277). A little later (*ibid*: 278), they note that:

Any discussion of the utility function for money must leave room for the effect of special circumstances on preferences. For example, the utility function of an individual who needs \$60,000 to purchase a house may reveal an exceptionally steep rise near the critical value.

Another important aspect of prospect theory is its emphasis on framing. Most discussion of this relates to the phenomenon where the same event can be framed as a gain or a loss, resulting (as prospect theory predicts) in different preferences from individuals (e.g., Tversky and Kahneman, 1986). But of course, for the same event (for the same person, at the same time) to be framed either as a loss or a gain, the subject would need to adopt distinct reference points as their comparison baseline; this again supports the claim that people can adopt different baselines than the one that Chung suggests. As such, the evidence in fact supports a rather different view than the one that Chung advocates. This is a view that we could still incorporate into a utilitarian framework; but it would no longer have the distributive implications Chung suggests.

These issues arise from the discrepancy between the somewhat objective threshold Chung discusses of ‘adequate functioning’, and the more subjective threshold bases derived from prospect theory. A purely subjective sufficientarian view may therefore face some of the same problems; but most sufficientarian views are not purely subjective. On the other hand, Chung might resolve this problem by dropping the reliance on prospect theory; but this would be a considerable alteration to the theory, and would leave its utilitarian credentials questionable. So while I cannot rule out some amendments to PU that will rescue it as a theory, this seems a fairly difficult task.

Conclusion

Chung (p.1933) offers a caveat at the end of his discussion, acknowledging that some may reject the assumptions about human psychology on which he relies. His response is that we may read his paper in this case as a refutation of utilitarianism, since it must rely on implausible assumptions to have egalitarian credentials. That conclusion may be too strong—one could reject egalitarianism—but this paper has shown that there are clear problems with the empirical bases of PU. I also suggested that even if the empirical evidence were there, the problems Chung raises for sufficientarianism are not so pressing as he supposes, and that PU has intuitive problems of its own.

References

- Benbaji, Yitzakh. 2005. The doctrine of sufficiency: A defence. *Utilitas* 17: 310–32.
- Bognar, Greg. 2015. Fair innings. *Bioethics* 29: 251–61.

- Chung, Hun. 2017. Prospect Utilitarianism: A better alternative to Sufficientarianism. *Philosophical Studies* 174: 1911–33.
- Crisp, Roger. 2003. Equality, priority and compassion. *Ethics* 113: 745–63.
- Dworkin, Ronald. 1981. Equality of what? Part I: Equality of welfare. *Philosophy and Public Affairs* 10: 185–246
- Frankfurt, Harry. 1987. Equality as a moral ideal. *Ethics* 98: 21–43.
- Goodin, Robert. 1999. Treating like cases alike, intergenerationally and internationally. *Policy Sciences* 32: 189–206.
- Herlitz, Anders and Nir Eyal. Forthcoming. Input and output in distributive theory. *Noûs*.
<https://doi.org/10.1111/nous.12392>
- Huseby, Robert. 2010. Sufficiency: Restated and defended. *The Journal of Political Philosophy* 18: 178–97.
- Huseby, Robert. 2020. Sufficiency and the threshold question. *The Journal of Ethics* 24: 207–23.
- McKerlie, Dennis. 2012. *Justice Between the Young and the Old* (Oxford: Oxford University Press).
- Mill, John Stuart. 1859/2003. On Liberty in ‘*Utilitarianism*’ and ‘*On Liberty*’ ed. by Mary Warnock. (Oxford: Blackwell), pp.88–180.
- Nielsen, Lasse. 2019. Sufficiency and satiable values. *Journal of Applied Philosophy* 36: 800–816.
- Roemer, John. 2004. Eclectic distributional ethics. *Politics, Philosophy & Economics* 3: 267–281.
- Shields, Liam. 2018. *Just Enough: Sufficiency as a Demand of Justice* (Edinburgh: Edinburgh University Press).
- Tversky, Amos, and Daniel Kahneman. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47: 263–91.
- Tversky, Amos, and Daniel Kahneman. 1986. Rational choice and the framing of decisions. *The Journal of Business*, 59: 251–78.