Input and output in distributive ethics

Abstract

Distributive theories evaluate distributions of goods based on individuals’ characteristics, e.g. how well off these individuals are, how deserving they are, and whether they fare below sufficiency. But such characteristics vary across possible worlds. Accordingly, distributive theories may differ as per the world by which they determine these characterizations. In particular: Should we fix the characteristics grounding distributive claims purely in the possible world that would take place absent any intervention on the distributing agent’s part (call it the “input” world)? Or, alternatively, should we assess each possible world that the distributor can bring about (call each an “output” world) purely by its own characteristics? We illustrate the importance of this distinction in relation to one distributive theory, prioritarianism. As we show, both input and output interpretations of prioritarianism are prominent in the literature, and there is something to be said for input prioritarianism. Ultimately, however, input prioritarianism is untenable.

Keywords: Counterfactuals, Distributive justice, Egalitarianism, Prioritarianism

I. Input vs. output prioritarianism
Prioritarianism is the view that "Benefiting people matters more the worse off these people are" (Parfit 1991: 19; Parfit 2012: 401). But how should one understand how badly off people are and, accordingly, what determines how much weight different benefits should be ascribed? Moral and political philosophers have explored what distributive domain, sphere, currency, chanciness, and time period define how badly off a person is for the purposes of prioritarian judgments. This talk expounds a new dimension: what possible world determines how badly off people are for these purposes? Parallel questions arise for distributive theories other than prioritarianism.

For some philosophers it may seem obvious that when prioritarians evaluate a world (or "outcome"/"situation"), what determines how much a benefit to some individual matters in it must be how well off she is in that particular world:

According to prioritarians... increased individual benefits have a successively decreasing impact on the overall goodness of a situation (Rabinowicz 2002: 6).

By ‘prioritarianism’, we mean the ranking of outcomes according to the sum of a strictly increasing and strictly concave transformation of individual well-being numbers (Adler & Holtug 2019: 3).¹

¹ See also Broome 1991: 199; Adler 2012: 360-362; and Adler & Treich 2017: 94.
On these definitions, the prioritarian goodness of an outcome depends only on how well off each individual is in that outcome. Only internal features of that outcome matter. So understood, prioritarianism satisfies the “permutation axiom” according to which, “for a given pattern of well-being, it should not matter which particular person is at which level” (Adler 2012: 52).

But this is not how prioritarianism is always understood, expressed, and operationalized. Consider, for instance, three philosophical bioethicists’ recent depiction of prioritarianism:

> We favor a moderate prioritarianism, which allocates benefits on the basis of a combined measure of the amount of benefit provided and the degree of disadvantage of the beneficiaries, so that a benefit is given greater weight the more disadvantaged the beneficiary would otherwise be (John, Millum & Wasserman 2017: 177—our italics).

Some of the same writers elsewhere suggest that, when applying prioritarianism to global health investment, “how badly off someone is depends on the actual outcomes that will befall her without intervention” (Sharp and Millum 2015: 113—our italics). This prioritarian strand seems to determine each individual’s level of disadvantage in a possible world in which no benefit has been distributed (at least not to that individual). That stance assesses outcomes based partly on features of other outcomes (in which the same individuals feature).
Let us ask, then: In which world is the prioritarian weight for the badly off determined? In each distribution being considered, or in a world without distributive intervention which serves as the “baseline” for assessing different distributions? This overlooked distinction affects both what prioritarianism recommends and what justifications are plausible for it.

On one interpretation, prioritarianism determines how badly off people are within each of the distributions under consideration. We shall call this the “output” interpretation of prioritarianism. This view states that, other things being equal, it is better to bring about distributions that are better for the worse off persons in these distributions. In other words, output prioritarianism evaluates each outcome based exclusively on the internal features of that outcome.

On a second interpretation, expressed and exemplified by John et al, priority for each candidate recipient is determined by the baseline of her standing in a world with no distributive intervention. That “given” or “natural” state is what determines candidates’ priority weights. We shall call this the “input” interpretation of prioritarianism. Input prioritarianism states that, other things equal, it is better to bring about distributions that are better for those who are worse off in a world in which the distributive intervention being considered does not take place. Candidate recipients’ priority weights remain robust across the possible worlds that different interventions could bring about.
Parallel distinctions pertain to other distributive theories. In desertarianism: Is who is (more) deserving determined in each world under consideration (output desertarianism), or only in possible worlds free from the distributive intervention (input desertarianism)? In sufficientarianism: Is whether people fare below sufficiency determined by how candidates fare in any distribution that the agent can bring about (output sufficientarianism), or only in the distribution that the agent encounters (input sufficientarianism)? In short, should distributive theories be outputty, or inputty? For presentational simplicity, we discuss this distinction between output and input views by reference to prioritarianism. However, we believe that the distinction between input and output views has relevance beyond prioritarianism, and one of the purposes of this talk is to draw attention to the importance of keeping distributive theories open about which possible worlds specify them.

II. **Sharpening the distinction**

Derek Parfit’s canonical description of prioritarianism can be read in two ways:

*Input prioritarianism*: Benefitting people always matters. And in each distribution considered, benefits to people matter more the worse off these people would be *if no (re)distributive intervention took place*.

*Output prioritarianism*: Benefitting people always matters. And in each distribution considered, benefits to people matter more the worse off these people would be *in that distribution*. 
Input prioritarianism is “baseline-dependent” since it fixes how much different benefits matter with reference to a particular baseline (how well off people are if nothing is done). Output prioritarianism is “baseline-independent”, and fixes how much a benefit matters only by the various worlds following distributive decisions.

The following case shows the different practical implications of these two interpretations of prioritarianism, and, hence, more generally, the importance of being clear on which possible worlds distributive criteria refer to:

*Case 1:*

<table>
<thead>
<tr>
<th></th>
<th>Do nothing</th>
<th>Do A</th>
<th>Do B</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Individual X</em></td>
<td>1</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td><em>Individual Y</em></td>
<td>6</td>
<td>9</td>
<td>7</td>
</tr>
</tbody>
</table>

Here, doing nothing is worse for everyone, so that option is ruled out; and whether the agent does A or B, the total amount of benefits is the same (they are 17 either way). Therefore, priority weights—and the possible world which determines these weights—make a decisive difference. On the output interpretation of prioritarianism, priority weights are fixed in each distribution considered. In deciding between Do A and Do B, what matters is only that the
worse off individual in each intervention is doing better under A (at 8) than under B (at 7).
Therefore, Do A is preferable overall. On the input interpretation of prioritarianism, since
*Individual X* is worse off than *Y* under Do nothing (which functions as a priority-weights-fixing
baseline), it matters more to benefit *X*, so Do B is preferable overall.

Consider an illustration of Case 1’s distributions:

Two newborn individuals have just been infected by the rare Pain Bug, which
gives all untreated individuals terrible lifelong pain. *More Exposed (Individual X)*
has been exposed to more of the bug, and her pain is so horrible that her
lifetime well-being would only be 1 (u=1) if she receives no treatment. *Less
Exposed (Individual Y)* has been exposed to less, and her pain would be less bad
(u=6) if she receives no treatment. Two drugs exist against Pain Bug but they are
prohibitively expensive. Fortunately, a donor is willing to purchase a single drug
type for both babies (cheaper per token this way), and for them alone, for life. So
the question arises which drug type, if any, the health agency should offer both
individuals.

Offering both individuals the conventional drug (Do A) would help both a lot but
the life of *More Exposed* (u=8) would remain worse than that of *Less Exposed*
(u=9). The conventional drug has only limited power against severe infections
like *More Exposed*’s.
The special drug uses the power of Pain Bug exposure against itself. So giving both individuals that drug (Do B) would, curiously, make the life of *More Exposed* (u=10) better than that of *Less Exposed* (u=7).

*Figure 1* below shows how input and output views would analyze this case. The contiguous black arrow shows the “input” world of nonintervention, which is the baseline for input prioritarianism. In this world, the distribution is (1;6). There are also two options for intervention, the outcomes of which are designated by broken lines. Intervention A leads to (8;9) and intervention B, to (10;7). Input prioritarians and output prioritarians both count positively the benefits to each recipient in each distribution. But they determine the weights assigned to those benefits differentially. Output prioritarians fix these weights with reference to how well off each person is following each intervention considered (flipping white arrows). Input prioritarians fix these weights with reference to the input world. They then apply these weights onto each intervention considered (dark grey arrows).
Besides these differences in practical implications, input and output prioritarianism are best supported by different philosophical foundations. For example, the oft-observed need to infuse

Figure 1: Case 1 analyzed by input prioritarians (dark grey priority weights) and by output prioritarians (flipping white priority weights), respectively.
utilitarianism (clearly an output view) with sensitivity to distributive considerations tends to
generate an output version of prioritarianism. The distributing agent’s alleged obligations
toward the people encountered would tend to support an input approach.

III. Taking input prioritarianism seriously

Many passages by prominent philosophers who discuss prioritarianism suggest that they have
input prioritarianism in mind:

Prioritarianism holds that institutions and practices should be set and actions
chosen to maximize moral value, with the stipulation that the moral value of
obtaining a benefit (avoiding a loss) for a person is greater, the greater the well-
being gain that the person would get from it (the smaller the loss in well-being),
and greater, the lower the person’s lifetime expectation of well-being prior to
receipt of the benefit (loss) (Arneson 2000: 343—our italics).

On the priority view the treatment for the very severe impairment will have a
higher expected moral value simply by virtue of the fact that the initial state
from which her utility would be increased would be at a lower absolute level
According to [prioritarianism] an increment in an individual’s well-being matters more the lower the level of well-being from which this increment takes place (Voorhoeve & Fleurbaey 2016: 932—our italics).²

[Prioritarianism] gives greater priority to the recipients who are starting off at lower baseline welfare levels... The prioritarian, under this new interpretation, wants to say that a benefit of a given absolute size is morally more important or matters more if it goes to someone whose baseline is lower. So the same +10 unit benefit would have greater moral importance if it were to go to someone whose baseline is low, but it would have smaller moral importance if it were to go to someone whose baseline is high (Feldman 2016: 165-166–our italics).

One’s claim to a benefit is stronger, it seems, the worse off one would be were it unsatisfied (Nebel 2017: 908 —our italics).

In these quotations, Arneson, Otsuka and Voorhoeve, Voorhoeve and Fleurbaey, Feldman, and Nebel all describe prioritarianism in a way that, to us, reads input prioritarian. The view they

² See also: “Increments in a person’s well-being have positive but diminishing marginal moral value— an increment that takes place from a lower level receives a higher “priority weight” than an increment that takes place from a higher level” (Voorhoeve & Fleurbaey 2016: 934—our italics).
describe seems to determine how much benefits to different individuals matter with reference to how the individuals fare when there is no intervention. Using different language, they all seem to ascribe particular importance to the baseline that we would call the input world: cf. “well-being prior to receipt of benefit”, “the initial state from which her utility would be increased”, “the level of well-being from which this increment takes place”, “lower welfare baseline levels” and “were it unsatisfied”.3

It may be tempting to dismiss these seeming expressions of input prioritarianism as mere careless writing. Indeed, Voorhoeve,4 Otsuka,5 and perhaps other authors we quote elsewhere

3 In principle, one could read some of these passages as expressions of what Adler calls the “claim-across-outcome” view, so that they do not refer to a baseline; instead they recommend a prioritarian pairwise comparison of outcomes, in which priority weights for individuals in each of two outcomes is established by the outcome among those two in which that individual fares worse (cf. Adler 2012: 331-337). But some of these quotations clearly define the baseline as no intervention, not as intervention that is worse for the relevant individual.

4 Voorhoeve elsewhere defines prioritarianism as “the view that each person’s welfare has diminishing marginal moral value and that the moral value of a person’s welfare depends only on that person’s level of welfare, and not on how anyone else fares” (Voorhoeve 2015: 201). This recalls the quotes from Rabinowicz and from Adler & Holtug.

5 Otsuka writes elsewhere: “All that matters, on the priority view, is how well off people are in absolute terms: the worse off they are, the more their well-being matters” (Otsuka 2015: 3—
present prioritarianism in its output version. We cannot rule out the possibility that some of the generally-careful philosophers we quote tried to convey output prioritarianism, infelicitously. But input interpretations of prioritarianism are expressed so frequently and, for some of these authors, repeatedly enough that the phenomenon seems deeper, and worth exploring. Furthermore, some of these authors offer inputty canonical statements, or suggest possible foundations for prioritarianism that support an input approach.\(^6\)

---

our italics). Otsuka’s present tense, the simplicity of his description, and his clarification that it captures all that matters preclude the relative complexity of input prioritarianism, in which the priority weight springs from an exogenous distribution. Likewise, in a more recent paper that does not directly address prioritarianism, Otsuka suggests that one of the things that unites “consequentialist theories, including utilitarianism and prioritarianism”, is that they consider “facts regarding how any particular individual would fare in the one distribution versus the other” entirely irrelevant (Otsuka 2017: 194).

\(^6\) Voorhoeve analysis elsewhere of the strength of competing claims reveals input-prioritarian commitment (Voorhoeve 2014). Nebel elsewhere defends the use of the status quo as a reference point for normative assessments (Nebel 2015). Feldman explicitly argues that prioritarianism should be understood as a theory about new distributions to individuals whose prioritarian weight is determined by the baseline levels prior to the distributions, and holds this is what Parfit had in mind (Feldman 2016: 156-166).
IV. Why input prioritarianism is untenable

The following arguments against input prioritarianism are together decisive: (1) Distributions matter even when they concern future people. (2) Input prioritarianism supports regressive redistributions that do not boost combined well-being. (3) Input prioritarianism demands pointless and even circular redistributions.

1. Distributions matter even when they are between future people

Sometimes, what distributors must assess is who would be worse off among people who would come into existence only if distributors intervene. Intuitively, distributive theories should remain informative on such occasions. Yet, input prioritarianism is completely uninformative. Consider:

Case 4:

A couple who wants twins requires in-vitro fertilization, and has four Pain-Bug infected eggs. Two were more exposed to the bug than the other two. Fertilizing any egg is sure to lead to a live birth, and no more eggs can be extracted or procured. The intended parents will not create children with horrible lives, so prohibitive drug prices permit the clinic to proceed only if a drug donor is willing to make drugs available to both twins. As the clinicians know, a drug donor is
willing to fund the same drug type to both twins for life, but she is a bit odd. She will decide whether to make the donation and which drug to donate based on who is created, in a complicated fashion that, to cut a long story short, leaves open the following two options only:

\[
\begin{array}{ccc}
\text{Create More Exposed &} & \text{Create More Exposed’ &} \\
\text{Less Exposed + give} & \text{Less Exposed’ + give} \\
\text{conventional drug} & \text{special drug}
\end{array}
\]

<table>
<thead>
<tr>
<th></th>
<th>Create More Exposed &amp;</th>
<th>Create More Exposed’ &amp;</th>
</tr>
</thead>
<tbody>
<tr>
<td>More Exposed</td>
<td>8</td>
<td>-</td>
</tr>
<tr>
<td>Less Exposed</td>
<td>9</td>
<td>-</td>
</tr>
<tr>
<td>More Exposed’</td>
<td>-</td>
<td>10</td>
</tr>
<tr>
<td>Less Exposed’</td>
<td>-</td>
<td>7</td>
</tr>
</tbody>
</table>

The literature rightly expects a prioritarian principle to favor the option to the left to the one to the right (Holtug 2010, 2012; Adler & Treich 2017). On the output view, it is clear enough how prioritarianism can do that. After all, the worse off person under that option fares best, and total well-being is no less than under the alternative.

On the input view, by contrast, how much the benefits to the various individuals matter depends on their well-being levels if nothing is done. Yet here, if nothing is done there would be no individuals so either there is no answer as to how much the different benefits matter, or
the two options are tied. Either way, whether input prioritarianism is silent here or indifferent between the two options, its implications are counterintuitive. The intuitive judgement is that the option on the left is preferable to the one to the right.

2. Input prioritarianism accepts regressive redistributions with no gain in combined well-being

Consider:

Case 5:

<table>
<thead>
<tr>
<th></th>
<th>Do nothing</th>
<th>Redistribute</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual X</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>Individual Y</td>
<td>9</td>
<td>10</td>
</tr>
</tbody>
</table>

Intuitively, prioritarianism opposes this redistribution, which increases inequality with no gain in overall well-being.

Output prioritarianism says exactly that. It evaluates alternative outcomes based only on their internal features, namely, the internal characteristics of (9;9) and (8;10) respectively. Since combined well-being is the same and the Do nothing distribution is better for the worse off in that distribution, output prioritarianism prefers it and opposes redistribution.
Input prioritarianism, on the other hand, is *indifferent* between staying at (9;9) and moving to (8;10). Combined well-being is the same either way, and under the baseline of (9;9) the individuals are as well off as one another. There is therefore no difference in the priority weights that, if it existed, might have counted against redistributing from (9;9) to (8;10).

3. **Input prioritarianism demands pointless or even circular redistributions**

Intuitively, it is pointless to redistribute so that the affected individuals completely switch position: the (formerly) worse off become as well off as the (formerly) better off were, and *vice versa*. Input prioritarianism, however, would recommend doing precisely that. It must ordain moving from e.g. (7;10) to (10;7) as long as the first outcome is identified as the input and there are no transaction costs. The reason is that both outcomes involve the same combined well-being and the second outcome is better for the worse off individual in the input world, which is the baseline that fixes priority weights according to input prioritarianism.

Indeed, input prioritarianism does not just recommend pointless redistributions, it recommends redistributions that are in a sense circular. Once (10;7) has been chosen, it is plausibly considered the input for future distributive decisions. That means that the move back from (10;7) to (7;10) is recommended by input prioritarianism as well, because it is better for those who are worse off under the new baseline. Once this decision is made, what counts as
the input for future decisions changes yet again, justifying yet another move from (7;10) to (10;7). And so on.

Besides being theoretically problematic, this circular ranking of outcomes creates two practical problems. First, to govern certain choice situations, some final decisions need to be laid down in advance, e.g. in relevant regulations, instructions, and decision-making computer programs. But such advance directives cannot provide determinate answers when the process would be circular *ad infinitum*.

Second, users of an input-prioritarian decision rule can be money-pumped. A committed input prioritarian will not only opt for (10;7) over (7;10). She will in fact be willing to take on a small cost, e.g. of 1 cent, in order to make that move. However, if what we just said is true, then once she has purchased this opportunity and the move takes place, she will favor going back, and again be willing to take on a small cost if that lets her change her decision from (10;7) to (7;10). And so on.

Output prioritarianism does not recommend any pointless circular, indeterminate, or money-pump-permitting evaluations. Instead, output prioritarianism deems the two outcomes—(7;10) and (10;7)—as equivalent, in line with commonsense intuition.\(^7\)

\(^7\) A possible answer to the circularity challenge may be that what counts as the source or input cannot alter once it is fixed for the first time. Once presented with (7;10), how well off
**Conclusion**

This talk introduced a new distinction between distributive theories. Using prioritarianism for illustration, it showed that distributive theories’ recommendations can vary based on whether they fix priorities with reference to the input world, which the distributing agent inherits; or, alternatively, exclusively with reference to the internal features of the output worlds, which the distributing agent can bring about through her distributive intervention. We made a case both for taking input prioritarianism seriously and for ultimately rejecting it.

---

individuals are and what priority they command is fixed, and it does not change even after the move to (10;7). That would have preempted the repeat cycles (not the initial pointless move). However, oftentimes this stipulation would seem unmotivated. Once agents change their minds and have every intention of implementing (10;7), or actually implement it, why not see the input for the next decision as (10;7)?
References


Otsuka, Michael. 2017. How it makes a difference that one is worse off than one could have been. *Politics, Philosophy and Economics* 17: 192-215.


