

# *The Ex-Ante Difference Principle Can Be Worse for Everyone Whatever Happens*

Johan E. Gustafsson\*

Draft: March 13, 2023 at 5:52 p.m.

ABSTRACT. In a recent paper, I argued that the *Ex-Ante* Difference Principle would not be chosen behind the Veil of Ignorance. In a response, Alexander Motchoulski and Phil Smolenski try to defend the *Ex-Ante* Difference Principle from this attack and claim that there are formal proofs that this principle would be chosen behind the Veil of Ignorance after all. In this paper, I rebut Motchoulski and Smolenski's objections and show that these alleged proofs don't work. In doing so, I present a counter-example that does not rely on any access to probabilities and a case where the *Ex-Ante* Difference Principle is worse for everyone whatever happens (making any worries about risk attitudes irrelevant).

According to John Rawls's version of Social Contract Theory, principles of justice are justified if they are the principles that individuals trying to secure their own ends would agree to in *the Original Position*, where they must deliberate behind the Veil of Ignorance.<sup>1</sup> Behind *the Veil of Ignorance*, the individuals do not know who they are and they cannot assign or estimate probabilities about who they are in society. Rawls argues that the individuals behind the Veil of Ignorance would agree to the Difference Principle, which demands that social and economic inequalities are 'to the greatest expected benefit of the least advantaged members of society'.<sup>2</sup> Rawls's formulation makes clear that he favours the *Ex-Ante* Difference Principle — a version of the Difference Principle that is concerned with the worst off in terms of expectations.<sup>3</sup>

In a recent paper, I argued that the *Ex-Ante* Difference Principle would not be chosen behind the Veil of Ignorance. In a response, Alexander

\* I would be grateful for any thoughts or comments on this paper, which can be sent to me at [johan.eric.gustafsson@gmail.com](mailto:johan.eric.gustafsson@gmail.com).

<sup>1</sup> Rawls 1971, pp. 136–142; 1999, pp. 118–123; 2001, pp. 85–89.

<sup>2</sup> Rawls 1974, p. 142.

<sup>3</sup> For an argument that Rawls defends the *ex-ante* version of the Difference Principle, see Gustafsson 2018, p. 597n29.

Motchoulski and Phil Smolenski defend the *Ex-Ante* Difference Principle from this attack and argue that there are formal proofs that this principle would be chosen behind the Veil of Ignorance after all. In this paper, I will rebut Motchoulski and Smolenski's objections (§§2–4) and show that these alleged proofs don't work (§5). In doing so, I present a counterexample that does not make use of any probabilities (§3) and a case where following the *Ex-Ante* Difference Principle would be worse for everyone whatever happens than following some alternative principle (§4).

\* \* \*

Before we start, however, some clarifications are needed. In Rawls's theory, the Difference Principle is subordinate to the Principle of Justice (demanding equal basic liberties), the Principle of Fair Equality of Opportunity (demanding public offices and social positions to be open to all), and the Just Savings Principle (demanding sufficient savings for the future).<sup>4</sup> To focus on the Difference Principle, we will assume that, in the cases we will discuss, these more prioritized principles are satisfied.

Moreover, to simplify the presentation, we will measure how well off people are in terms of well-being rather than Rawls's preferred currency of primary goods. (But, if we wanted, we could replace well-being levels throughout with indexes of primary goods.<sup>5</sup>) Hence we state the *Ex-Ante* Difference Principle as follows:<sup>6</sup>

*The Ex-Ante Difference Principle* Let the social value of a prospect be equal to the minimum expected well-being of any person in the prospect. Choose a prospect with a maximal social value among all alternative prospects.

This principle is only intended to be applied to choices between alternative basic structures of society. The basic structure of society is how fundamental rights and duties are distributed by major social institutions and

<sup>4</sup> Rawls 1971, pp. 302–303; 1999, pp. 266–267; 2001, p. 61.

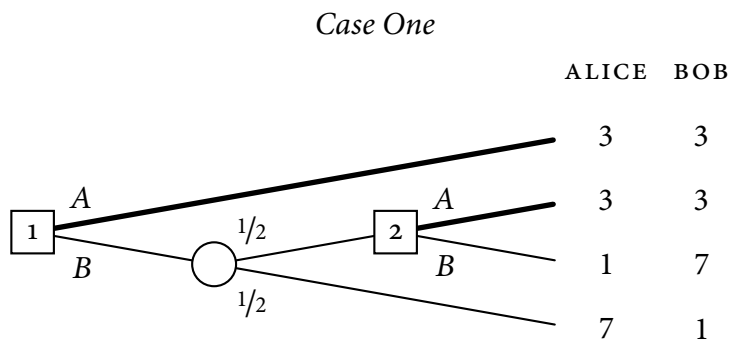
<sup>5</sup> This was pointed out in Gustafsson 2018, pp. 590–591. This replacement blocks Chung's (2021, pp. 450–455) objection that relies on distinguishing well-being levels and indexes of primary goods. I am not denying that well-being is distinct from primary goods. The point is not that the difference doesn't matter. The point is that we can replace well-being levels with indexes of primary goods in each of the principles and cases we are going to discuss, and then run the argument again (with the same result).

<sup>6</sup> This formulation comes from Gustafsson 2018, p. 597.

how these institutions determine the distribution of advantages from social cooperation.<sup>7</sup>

### 1. The risk-neutral case

Let us first consider a simplified version of my original counter-example to the *Ex-Ante* Difference Principle. This example makes use of the simplifying assumption that the parties behind the Veil of Ignorance are risk neutral — an assumption that (as shall see later) can be dropped. Let *A* and *B* be two alternative basic structures of society:<sup>8</sup>



In this diagram, the boxes represent choice nodes, where we have a choice between *A* and *B*. And the circle represents a chance node, where there is a one-in-two probability that chance goes up and a one-in-two probability that chance goes down. We assume that these probabilities have an objective basis. If we choose *B* at the initial choice node 1, we reach the chance node. Then, if chance goes up, we reach choice node 2.

At node 2, the outcome of *A* gives each of Alice and Bob a well-being of 3 and the outcome of *B* gives Alice a well-being of 1. Hence the *Ex-Ante* Difference Principle prescribes *A* at node 2.

Using backward induction, we take this knowledge of what would be chosen at node 2 into account at node 1.<sup>9</sup> Given that *A* would be chosen

<sup>7</sup> Rawls 1971, p. 7; 1999, p. 6.

<sup>8</sup> This is a simplified, cleaned up version of a case in Gustafsson 2018, p. 598.

<sup>9</sup> See Selten 1975 and Rosenthal 1981, pp. 94–95. Note that this case (like the others we shall consider) is *BI-terminating*, that is, the choices that are prescribed by backward induction in this case are final in the sense that they do not lead to any further choices; see Rabinowicz 1998, p. 101. In *BI-terminating* cases, we can defend the choices prescribed by backward induction in a more compelling way than in other kinds of cases; see Rabinowicz 1998, pp. 118–121 and Broome and Rabinowicz 1999, pp. 240–241. Note also

at node 2, choosing *B* at node 1 results in a fifty-fifty gamble for Bob between getting a well-being of 1 or 3. Hence choosing *B* at node 1 gives Bob an expected well-being of 2. But choosing *A* at node 1 gives everyone an expected well-being of 3. Hence the *Ex-Ante* Difference Principle prescribes *A* at node 1.

But, if we choose *B* at node 1 and stick with *B* if we reach node 2, everyone faces a fifty-fifty gamble between getting and a well-being of 1 or 7 so that everyone gets an expected well-being of 4. Hence everyone gets an expected well-being of 4 if we choose and stick with *B*, but, if we follow the prescriptions of the *Ex-Ante* Difference Principle, everyone merely gets an expected well-being of 3. Hence everyone gets a worse expectation if we follow the *Ex-Ante* Difference Principle in Case One than if we violate it throughout. So, no matter who they are in society, the parties know that the *Ex-Ante* Difference Principle would be worse for them than some alternative principle. So, if the parties are risk neutral, they wouldn't choose the *Ex-Ante* Difference Principle.

## 2. Motchoulski and Smolenski's objections

Case One shows that the *Ex-Ante* Difference Principle violates the following principle:<sup>10</sup>

*The Weak Sequential Ex-Ante Pareto Principle* If each person has a higher expected well-being in prospect *x* than in prospect *y*, then a plan whose outcome is *y* is not followed if there is an alternative plan available whose outcome is *x*.

Motchoulski and Smolenski object that this principle is a principle for *collective* choice just like the Difference Principle.<sup>11</sup> Hence to assume the Weak Sequential *Ex-Ante* Pareto Principle in a critique of the *Ex-Ante* Difference Principle is to assume a point at issue.

This objection misses the mark, however. As should be clear from the presentation in section 1, the Weak Sequential *Ex-Ante* Pareto Principle is never assumed in the argument. It's true that the *Ex-Ante* Difference

---

that, while backward induction is a normative assumption, it is an assumption about individual rationality rather than collective rationality.

<sup>10</sup> Gustafsson 2018, p. 599.

<sup>11</sup> Motchoulski and Smolenski 2019, p. 681. See also Chung 2021, pp. 455–463 for a similar objection.

Principle violates the Weak Sequential *Ex-Ante* Pareto Principle in Case One. But the reason the parties behind the Veil of Ignorance would reject the *Ex Ante* Difference Principle is not that it violates Weak Sequential *Ex Ante* Pareto Principle. They would reject it because they see that there are situations in which choosing the *Ex Ante* Difference Principle makes each one of them (no matter who they are) worse off in expectation than if they chose another principle. Whenever a principle violates the Weak Sequential *Ex-Ante* Pareto Principle, everyone behind the Veil of Ignorance knows that their interests would be better served in that case if they chose some alternative principle. Crucially, they can figure this out without access to any probabilities about who they are in society, since *everyone* has a lower expectation if the Weak Sequential *Ex-Ante* Pareto Principle is violated. But the parties don't need to accept the Weak Sequential *Ex-Ante* Pareto Principle to figure this out.

Motchoulski and Smolenski also object to the assumption of risk neutrality in Case One. They object that Rawls only assumed that the chosen principles don't depend on any special attitudes towards risk.<sup>12</sup>

A first response is that, if we don't assume that the parties to be risk neutral, it is hard to see why they would favour the *Ex-Ante* Difference Principle. The *Ex-Ante* Difference Principle, after all, focuses on the expected well-being of the worst off (that is, the worst off in terms of expected well-being). If the parties aren't risk neutral, it seems that they would have no reason to attach any significance to *expectations* of well-being. So we have the following fork: *either* we assume risk neutrality and then Case One can't be dismissed *or* we do not assume risk neutrality and then the parties wouldn't attach any significance to expectations. Thus in neither case would the parties agree to the *Ex-Ante* Difference Principle. A second response is that we can modify the case so that it doesn't rely on the assumption of risk neutrality, as we shall see in sections 3 and 4.<sup>13</sup>

<sup>12</sup> Motchoulski and Smolenski 2019, pp. 683–684, citing Rawls 1999, p. 148. Motchoulski and Smolenski ignore, however, the appendix to Gustafsson 2018, p. 604 which shows how the case can be modified so that it only requires that the parties lack extreme attitudes towards risk.

<sup>13</sup> Motchoulski and Smolenski (2019, pp. 685–686) complain that, if we assume that the parties are risk neutral, then it follows by Harsanyi's (1955) theorem that the parties would agree a utilitarian principle. But Harsanyi's proof requires more than just risk neutrality of individuals; it also requires risk neutrality for the social ordering, which is not assumed here. And Harsanyi's (1953) earlier veil-of-ignorance proof assumes that parties have an equal probability of being anyone in society, which is not assumed here. I am puzzled by Motchoulski and Smolenski's (2019, p. 686) claim, regarding Harsanyi's

Motchoulski and Smolenski next object that Rawls rules out the parties' having any basis for assigning probabilities to outcomes.<sup>14</sup> They cite Rawls's notorious no-probabilities passage:<sup>15</sup>

[T]he veil of ignorance excludes all knowledge of likelihoods. The parties have no basis for determining the probable nature of their society, of their place in it. Thus, they have no basis for probability calculations.

Yet blocking probabilities doesn't help the case for the *Ex-Ante* Difference Principle. If the parties can't consider probabilities at all, it's hard to see how they could properly assess the *Ex-Ante* Difference Principle, since that principle deals with probabilities. If one can't consider probabilities, one cannot, in general, assess the *expectations* of well-being for the worst off, which is crucial for the *Ex-Ante* Difference Principle. The principles of justice need to address risks and their probabilities in society. It makes no sense to choose these principles without being able to consider probabilities.

Moreover, contrary to what may be suggested by the no-probabilities passage, Rawls did not, I think, intend to block all probability calculations. In his discussion of the Principle of Sufficient Reason, Rawls is careful to only rule out probabilities that aren't based on objective facts:

I shall assume, . . . , to fill out the description of the original position, that the parties ignore estimates of likelihoods not supported by particular facts and that derive from the principle of insufficient reason.<sup>16</sup>

And he describes the parties in the Original Position as

---

result, that 'a Rawlsian need not be concerned that a utilitarian model leads to a utilitarian outcome, since among the aims of the [Original Position] is to articulate a shared moral point of view from which we can critique utilitarianism.' The point of the Original Position is to find the principles of justice, regardless of whether they turn out to be utilitarian. And, if Harsanyi's proof shows that utilitarianism follows when standard rational-choice theory is applied to both individuals and the social planner (given some further assumptions Rawlsians accept), then that should be worrying to Rawlsians. Even if you currently believe  $p$ , you shouldn't ignore evidence that  $p$  is false.

<sup>14</sup> Motchoulski and Smolenski 2019, p. 684.

<sup>15</sup> Rawls 1999, p. 134. The corresponding passage is, notably, less extreme in Rawls 1971, p. 155.

<sup>16</sup> Rawls 1999, p. 150. Rawls 1971, p. 173 has a slightly different wording.

taking effective means to ends with unified expectations and objective interpretation of probability<sup>17</sup>

These passages make little sense if Rawls intended to rule out all probability calculations in the Original Position, rather than merely estimates with a subjective basis, derived from the Principle of Sufficient Reason. It might help to remember Rawls's motivation for blocking any probability calculations at all. The reason for blocking estimates of probabilities about one's identity in society is that these estimated probabilities would estimate the information that the Veil of Ignorance is supposed to block to ensure impartiality.<sup>18</sup> This motivation does not extend to blocking all probabilities about risks in society.

The most charitable interpretation of the no-probabilities passage is that Rawls only meant to exclude knowledge and estimates of likelihoods about what society the parties *actually live in*.<sup>19</sup> In this way, the parties are unable to estimate whether they are, for example, rich or poor, and they have no way to choose a principle which is biased in favour of their actual situation or role in society. This still allows the parties to consider lots of potential situations they might be in and consider risks and probabilities based on objective facts in those potential situations. And, given that the probabilities in Case One have an objective basis, the parties can consider that case. Note, furthermore, that, even if the parties knew that Case One described their actual society, they still wouldn't gain any hints about their identity if they had access the probabilities for the chance node.<sup>20</sup> Hence, in the Original Position, there's no plausible basis for discarding considerations of probabilities like those in Case One. Nevertheless, as we shall see next, the probabilities in Case One are actually superfluous.

Finally, Motchoulski and Smolenski object to the sequential form of Case One.<sup>21</sup> Why, they ask, should we calculate expectations from node 1 rather than a some node at another time? To privilege a node at a certain time seems to admit time-sensitive information, which Rawls was care-

<sup>17</sup> Rawls 1971, p. 146; 1999, p. 127.

<sup>18</sup> Rawls 1971, p. 171; 1999, p. 147.

<sup>19</sup> See Gustafsson 2018, pp. 595–596n28 for a more extensive defence of this interpretation.

<sup>20</sup> Allowing probabilities about potential societies based on objective facts does not amount to a situation where 'the veil has been partially lifted', as Motchoulski and Smolenski (2019, p. 684) suggest. The point is that the Veil of Ignorance, properly understood, never hid those probabilities.

<sup>21</sup> Motchoulski and Smolenski 2019, p. 686.

ful to rule out.<sup>22</sup> We don't want the parties to choose principles that are biased in favour of the interests at a certain time.

But note first that we would have the same problem if we only considered synchronic situations. Why wouldn't the parties pick a principle that is biased in favour of the interests at the time of the synchronic choice?

Note also that the sequential application of the principles of justice is unavoidable in light of the following two desiderata: First, the parties should be concerned with lifetime well-being rather than well-being at a time.<sup>23</sup> Second, during lifetimes, the basic structure of society may need to be adjusted in accordance with the principles of justice to address new injustices that may arise due to changing circumstances such as new technologies and changes in culture.<sup>24</sup> As the parties lifetime well-being may depend on multiple applications of the principles of justice to choose and revise the basic structure of society, they need to take sequential choices into account.

But how do we avoid a bias toward the interests at a certain time if we take sequential (or synchronic) choices into account? The trick is that the parties in behind the Veil of Ignorance don't know which sequential situation they are in. So they need to consider all kinds of potential situations. If the parties find a situation they may potentially be in where it's guaranteed, no matter who they are, that the *Ex-Ante* Difference Principle would be worse for them than some alternative principle, then they won't choose the *Ex-Ante* Difference Principle.

### **3. A case with neither probabilities nor risk neutrality**

The assumption that the parties are risk neutral and have access to probabilities in Case One is inessential. It suffices that the parties' risk attitudes are non-extreme in the sense that there are some well-being levels  $\alpha > \beta > \gamma$  such that a gamble between getting a well-being of  $\alpha$  or  $\gamma$  is better than a certainty of getting a well-being of  $\beta$ . This is plausible as we may pick an  $\alpha$  arbitrarily greater than  $\beta$  and a  $\gamma$  minimally lesser than  $\beta$ . Hence the potential gain from getting the gamble between  $\alpha$  and  $\gamma$  rather than getting  $\beta$  is arbitrarily large and the potential cost is arbitrarily small.

Now, consider

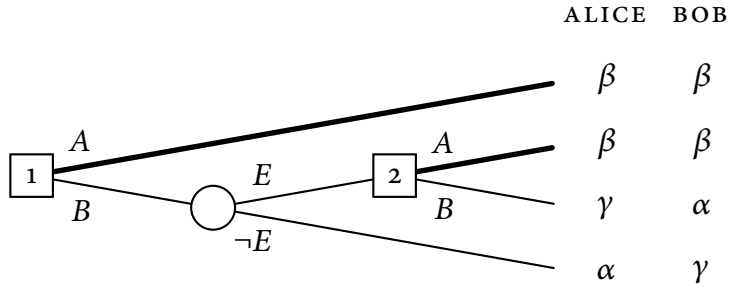
<sup>22</sup> Rawls 1971, p. 139; 1999, p. 120.

<sup>23</sup> Rawls 1971, p. 64; 1999, p. 56; 2001, p. 59. The focus on lifetime well-being is needed to avoid the sequential problem in Haslett 1985, pp. 111-112.

<sup>24</sup> Rawls 1977, p. 164; 1993, p. 284.



## Case Two



This case has the same structure as Case One, but in this case we don't have access to the probabilities in the chance node, which depends on chance event  $E$ . We only have that it is both possible that  $E$  happens and that  $E$  does not happen. Hence we have that it is both possible that chance goes up ( $E$  happens) and possible that chance goes down ( $E$  does not happen).

At node 2, if  $A$  is chosen, everyone gets a well-being of  $\beta$ , but, if  $B$  is chosen, Alice merely gets a well-being of  $\gamma$ . Since  $\beta > \gamma$ , the *Ex-Ante* Difference Principle prescribes  $A$  at node 2.

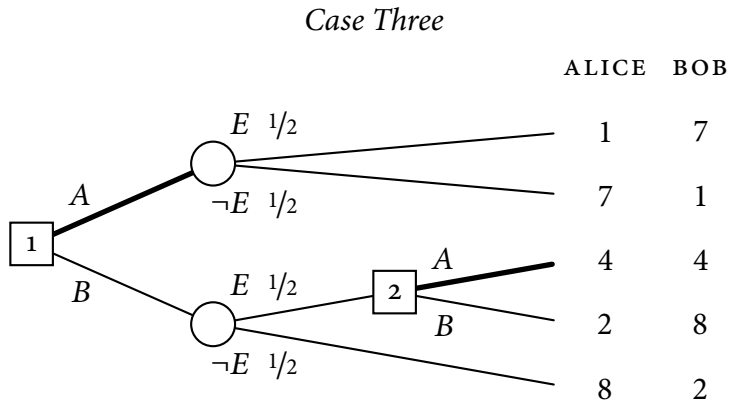
Using backward induction, we take this into account at node 1. If  $A$  is chosen at node 1, everyone gets a well-being of  $\beta$ . If  $B$  is chosen at node 1, Bob faces a gamble between  $\beta$  and  $\gamma$ . Since  $\beta > \gamma$ , the prospect of certainly getting a well-being of  $\beta$  stochastically and statewise dominates a gamble between getting a well-being of  $\beta$  if  $E$  happens or getting a well-being of  $\gamma$  if  $E$  does not happen. So the prospect Bob faces if  $B$  is chosen at node 1 is worse than the prospect everyone faces if  $A$  is chosen at node 1. Thus the *Ex-Ante* Difference Principle prescribes  $A$  at node 1.

If we choose  $B$  throughout, everyone faces a gamble between getting a well-being of  $\alpha$  or getting a well-being of  $\gamma$ . And, as we have assumed, that prospect is better than getting a well-being of  $\beta$  with certainty. Hence, in Case Two, it would be worse for everyone if the *Ex-Ante* Difference Principle were followed than if it were violated throughout.

For Case Two, we only assumed that the parties have a non-extreme risk attitude and we did not rely on any probabilities. Yet we can do better still.

#### 4. A case where the *Ex-Ante* Difference Principle is worse for everyone whatever happens

We may, in fact, sidestep any worries about risk attitudes. There are cases where the *Ex-Ante* Difference Principle is worse for everyone *whatever happens*. Consider



Here, there are two separate chance nodes. But they both depend on the same event  $E$ , which has a one-in-two chance of happening.

At node 2, the *Ex-Ante* Difference Principle prescribes  $A$ , since each of Alice and Bob then gets a higher well-being in the outcome of  $A$  than Alice gets in the outcome of  $B$ .

Using backward induction, we take this into account at node 1. The outcome of  $B$  at node 1 is then a prospect with a one-in-two chance of Alice getting a well-being of 8 and Bob getting a well-being of 2 and a one-in-two chance of Alice and Bob each getting a well-being of 4. So choosing  $B$  at node 1 gives Alice an expected well-being of 6 and Bob an expected well-being of 3. Choosing  $A$  at node 1, on the other hand, gives each of Alice and Bob a fifty-fifty gamble between getting a well-being of 1 or 7. So choosing  $A$  at node 1 gives everyone an expected well-being of 4. Hence the *Ex-Ante* Difference Principle prescribes  $A$  at node 1.

But compare the outcome of the *Ex-Ante* Difference Principle's prescriptions in Case Three — that is, choosing  $A$  throughout — with the outcome of the opposite of its prescriptions — that is, choosing  $B$  throughout:

	<i>E</i> happens		<i>E</i> does not happen	
	ALICE	BOB	ALICE	BOB
<i>A</i> at node 1	7	1	1	7
<i>B</i> at nodes 1 and 2	8	2	2	8

Regardless of whether event *E* happens, everyone gets a lower well-being level if *A* is chosen than if *B* is chosen throughout. Hence following the prescriptions of the *Ex-Ante* Difference Principle in Case Three is guaranteed to be worse for everyone than following the opposite prescriptions, whatever happens. The outcome of the *Ex-Ante* Difference Principle is not only dominated in expectations for everyone but also stochastically dominated and statewise dominated for everyone.<sup>25</sup>

So, no matter what risk-attitude the parties have and no matter who they are in society, they know that it would be worse for them if the *Ex-Ante* Difference Principle were followed than if the opposite of its prescriptions were followed. So the parties behind the Veil of Ignorance would not choose the *Ex-Ante* Difference Principle. For this argument, we only relied on probabilities when we applied the *Ex-Ante* Difference Principle — not when we assessed the prescriptions.

## 5. Two alleged proofs that the Difference Principle would be chosen

In their concluding remarks, Motchoulski and Smolenski claim that there are two formal proofs that the Difference Principle is entailed by the design features of the Original Position.<sup>26</sup>

The first is due to Amartya Sen. Motchoulski and Smolenski claim that Sen has shown that ‘one formalization of reciprocity, represented by Patrick Suppes’s grading principle of justice, directly entails the difference principle under conditions described in the [Original Position].’ This, however, is a misunderstanding of Sen’s proof, which shows something else. According to

<sup>25</sup> Stochastic Dominance should be compelling even if you are risk averse. Buchak (2013, pp. 37–38), for example, accepts both Statewise and Stochastic Dominance as requirements of rationality even though she defends risk-aversion. And, while Stochastic Dominance is a normative principle (that is, the rational requirement that one avoids choosing stochastically dominated prospects), it is a normative principle for individual rationality rather than collective rationality.

<sup>26</sup> Motchoulski and Smolenski 2019, pp. 289–290.

*The Grading Principle of Justice* Distribution  $x$  is more just than distribution  $y$  if and only if there is a one-to-one transformation from the set of individuals to itself such that one would prefer to be in the position of someone in  $x$  rather than in the position of the corresponding person in  $y$ , and also would prefer to be, or would be indifferent to being, in the position of *each* person in  $x$  than to be in the position of the corresponding person in  $y$ .<sup>27</sup>

What Sen actually proves is that, if  $x$  is more just than  $y$  according to the Grading Principle of Justice, then  $x$  is at least as just as  $y$  according to the Difference Principle.<sup>28</sup> This, however, does not show that the Grading Principle of Justice entails (or supports) the Difference Principle. First, note that it is equally true that, if  $x$  is more just than  $y$  according to the Grading Principle of Justice, then  $x$  is more just than  $y$  according to a utilitarian principle of justice. Second, and more crucially, the Grading Principle of Justice entails that the Difference Principle is false. To see this, note that the Difference Principle entails that the distribution <Alice at 2, Bob at 2> is more just than the distribution <Alice at 3, Bob at 1> but the Grading Principle of Justice entails that <Alice at 2, Bob at 2> is *not* more just than <Alice at 3, Bob at 1>.<sup>29</sup> So, if we accept the Grading Principle of Justice, we must reject the Difference Principle.

The second alleged proof is due to Steven Strasnick. The main premise in his derivation of the Difference Principle is

*The Priority Principle* For all individuals  $I$  and  $J$  and distributions  $x, y, u, v$ , if  $I$  has the same well-being in  $y$  as  $J$  has in  $v$ , then  $I$ 's preference for  $x$  over  $y$  shall have the same priority as  $J$ 's preference for  $u$  over  $v$ .<sup>30</sup>

Why think that the Priority Principle follows from the design features of the Original Position? It is not immediately clear what is meant by priority here, but it makes sense that each person's preferences should be given the same priority (in some sense) in the Original Position.<sup>31</sup> But then we may, just as well, accept

<sup>27</sup> Sen 1970, pp. 153,156. This principle is a generalization of a similar principle of two-person cases, which was proposed by Suppes (1966, pp. 288–293).

<sup>28</sup> Sen 1970, pp. 157–158.

<sup>29</sup> Sen (1970, p. 158) provides much the same counterexample.

<sup>30</sup> Strasnick 1976, p. 89.

<sup>31</sup> Strasnick's (1976, pp. 88–89) own argument for the Priority Principle is not compelling. See the objections in Goldman 1976, pp. 847–849 and Wolff 1976, pp. 852–854.

*The Converse Priority Principle* For all individuals  $I$  and  $J$  and distributions  $x, y, u, v$ , if  $I$  has the same well-being in  $x$  as  $J$  has in  $u$ , then  $I$ 's preference for  $x$  over  $y$  shall have the same priority as  $J$ 's preference for  $u$  over  $v$ .

And then, if Strasnick's proof is valid, we should be able to derive a *maxim* rather than *maximin* principle of justice, since his other assumptions are neutral between better and worse.<sup>32</sup>

Strasnick's derivation of the Difference Principle starts with a derivation of the following lemma:<sup>33</sup>

*Cancellation of Opposing Preference* In two person cases where one individual prefers  $x$  over  $y$  and the other individual prefers  $y$  over  $x$ , if these preferences have the same priority, then  $x$  and  $y$  are equally just.

If we grant that there is a relevant sense of priority is such that Cancellation of Opposing Preference holds, then the Priority Principle is implausible. While it makes sense to give the same priority or weight to each person's preference other things being equal to ensure fairness, it doesn't make sense to ignore the strength of these preferences. It doesn't seem unfair to give more priority to preferences with greater strength, given that the same priority is given to each person's preference if they have the same strength. But that is ruled out by Cancellation of Opposing Preference. This lemma entails, for example, that the distribution <Alice at 3, Bob at 1> is equally as just as <Alice at 1, Bob at 2>. This conflicts with Rawls's favoured Leximin version of the Difference Principle, which yields that <Alice at 3, Bob at 1> is more just than <Alice at 1, Bob at 2>.<sup>34</sup> So Rawls couldn't consistently accept the Priority Principle.

A final worry about the Priority Principle is that it's a principle about how to prioritize with respect to justice.<sup>35</sup> If this principle is built in to the Original Position, the resulting variation of Social Contract Theory loses much of its appeal. The compelling idea behind Social Contract Theory is that we get the principles of justice from what parties trying to secure

<sup>32</sup> The validity of Strasnick's proof has been challenged in Wolff 1976, pp. 857–858.

<sup>33</sup> Strasnick 1976, p. 92

<sup>34</sup> Rawls 1971, pp. 82–83.

<sup>35</sup> Strasnick (1976, p. 99) admits that his version of the Original Position is no longer value neutral.

their own interest in an initial situation of equality. That is, principles of justice are the output rather than the input of the social-contract set-up.

Summing up this section, Motchoulski and Smolenski claim that there are two formal proofs that the Difference Principle would be chosen behind the Veil of Ignorance. Neither works. So Motchoulski and Smolenski's positive case that the Difference Principle would be chosen doesn't work.

I wish to thank Jacob Barrett, Todd Nugent Karhu, Jakob Nebel, Dean Spears, Martin Peterson, Wlodek Rabinowicz, Luke Townsend, and Robert Paul Wolff for valuable comments.

## References

- Broome, John and Wlodek Rabinowicz (1999) 'Backwards Induction in the Centipede Game', *Analysis* 59 (4): 237–242.
- Buchak, Lara (2013) *Risk and Rationality*, Oxford: Oxford University Press.
- Chung, Hun (2021) 'On Choosing the Difference Principle behind the Veil of Ignorance: A Reply to Gustafsson', *The Journal of Philosophy* 118 (8): 450–463.
- Goldman, Alan H. (1976) 'Rawls's Original Position and the Difference Principle', *The Journal of Philosophy* 73 (21): 845–849.
- Gustafsson, Johan E. (2018) 'The Difference Principle Would Not Be Chosen behind the Veil of Ignorance', *The Journal of Philosophy* 115 (11): 588–604.
- Harsanyi, John C. (1953) 'Cardinal Utility in Welfare Economics and in the Theory of Risk-Taking', *The Journal of Political Economy* 61 (5): 434–435.
- (1955) 'Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility', *The Journal of Political Economy* 63 (4): 309–321.
- Haslett, D. W. (1985) 'Does the Difference Principle Really Favour the Worst Off?', *Mind* 94 (373): 111–115.
- Motchoulski, Alexander and Phil Smolenski (2019) 'Principles of Collective Choice and Constraints of Fairness: Why the Difference Principle Would Be Chosen behind the Veil of Ignorance', *The Journal of Philosophy* 116 (12): 678–690.

- Rabinowicz, Wlodek (1998) 'Grappling with the Centipede: Defence of Backward Induction for BI-Terminating Games', *Economics and Philosophy* 14 (1): 95–126.
- Rawls, John (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press.
- (1974) 'Some Reasons for the Maximin Criterion', *American Economic Review* 64 (2): 141–146.
- (1977) 'The Basic Structure as Subject', *American Philosophical Quarterly* 14 (2): 159–165.
- (1993) *Political Liberalism*, New York: Columbia University Press.
- (1999) *A Theory of Justice*, Cambridge, MA: Harvard University Press, revised edn.
- (2001) *Justice as Fairness: A Restatement*, ed. Erin Kelly, Cambridge, MA: Harvard University Press.
- Rosenthal, Robert W. (1981) 'Games of Perfect Information, Predatory Pricing and the Chain-Store Paradox', *Journal of Economic Theory* 25 (1): 92–100.
- Selten, R. (1975) 'Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games', *International Journal of Game Theory* 4 (1): 25–55.
- Sen, Amartya K. (1970) *Collective Choice and Social Welfare*, San Francisco: Holden-Day.
- Strasnick, Steven (1976) 'Social Choice and the Derivation of Rawls's Difference Principle', *The Journal of Philosophy* 73 (4): 85–99.
- Suppes, Patrick (1966) 'Some Formal Models of Grading Principles', *Synthese* 16 (3–4): 284–306.
- Wolff, Robert Paul (1976) 'On Strasnick's "Derivation" of Rawls's "Difference Principle"', *The Journal of Philosophy* 73 (21): 849–858.