[Published in *The Journal of Philosophy* 115 (11): 588–604, 2018. https://dx.doi.org/10.5840/jphil20181151134]

The Difference Principle Would Not Be Chosen behind the Veil of Ignorance*

Johan E. Gustafsson†

John Rawls argues that the Difference Principle (also known as the Maximin Equity Criterion) would be chosen by parties trying to advance their individual interests behind the Veil of Ignorance. Behind this veil, the parties do not know who they are and they are unable to assign or estimate probabilities to their turning out to be any particular person in society. Much discussion of Rawls's argument concerns whether he can plausibly rule out the parties' having access to probabilities about who they are. Nevertheless, I argue that, even if the parties lacked access to probabilities about who they are in society, they would still reject the Difference Principle. I argue that there are cases where it is still clear to the parties that it is not in any of their individual interests that the Difference Principle be adopted.

What, if anything, could justify a principle of social justice? One answer, from the political thought of the Enlightenment, is a social contract. According to *Social Contract Theory*, a principle of justice is justified if and only if it would be agreed to by parties trying to advance their individual interests in a certain initial situation.

In John Rawls's version of Social Contract Theory, this initial situation is *the Original Position*—an initial situation where the parties are situated behind a *Veil of Ignorance*. Behind this veil, the parties do not know who they are and they are unable to assign or estimate probabilities to their turning out to be any particular person in society.[1] Rawls argues that,

* I wish to thank Arif Ahmed, Krister Bykvist, Stephen Holland, Christopher Jay, Mary Leng, Martin O'Neill, Martin Peterson, Christian Piller, Wlodek Rabinowicz, Alan Thomas, and an anonymous referee for valuable comments.

† I would be grateful for any thoughts or comments on this paper, which can be sent to me at *johan.eric.gustafsson@gmail.com*.

[1] John Rawls, *A Theory of Justice* (Cambridge, MA: Harvard, 1971), pp. 136–42, and (Cambridge, MA: Harvard, 1999), pp. 118–23. Hereinafter referred to as TJ. See also Rawls, *Justice as Fairness: A Restatement*, ed. Erin Kelly (Cambridge, MA: Harvard, 2001), pp. 85–89. Hereinafter referred to as JF. A thinner veil, behind which one has an equal probability of turning out to be anyone, was first put forward in William Vickrey, "Measuring Marginal Utility by Reactions to Risk," *Econometrica*, x111, 4 (October 1945): 319–33, at p. 329.

behind the Veil of Ignorance, the parties would choose the Difference Principle (also known as the Maximin Equity Criterion).[2]

Much discussion of Rawls's argument concerns whether he can plausibly rule out the parties' having access to probabilities about who they are.[3] If the parties assign an equal probability to their turning out to be anyone in society, they would realize that they maximize their **[p. 589]** expected well-being if they agree to the Principle of Average Utility, rather than the Difference Principle.[4]

In this paper, I shall argue that, even if the parties lacked access to probabilities about who they are in society, they would still reject the Difference Principle. I shall argue that—even without assigning or estimating probabilities to their turning out to be any particular person in society—there are still cases where it is clear to the parties that it is not in their individual interests that the Difference Principle be adopted. Hence, behind the Veil of Ignorance, the parties would not choose the Difference Principle.

* * *

Before we begin, however, we should clarify some terminology. Following Rawls, we make a distinction between cases of *risk*, where there is an objective basis for estimating probabilities, and cases of *uncertainty*, where there is no such basis.[5] Moreover, we distinguish the Difference Principle from the following principle for choice under uncertainty:

> *The Maximin Rule for Choice under Uncertainty*
> Let the value of a prospect be equal to the worst possible final outcome of the prospect. Choose a prospect with a maximal value among all alternative prospects.[6]

[2] TJ (1971), pp. 118–92, (1999) pp. 102–67; and JF, pp. 80–134. Amartya K. Sen, *Collective Choice and Social Welfare* (San Francisco: Holden-Day, 1970), pp. 137, 157, put forward the first exact formulation of the Maximin Equity Criterion, based on some remarks in John Rawls, "Justice as Fairness," this JOURNAL, LIV, 22 (October 1957): 653–62, at p. 656; and John Rawls, "Distributive Justice," in Peter Laslett and W. G. Runciman, eds., *Philosophy, Politics, and Society: Third Series* (Oxford: Blackwell, 1967), pp. 58–82, at pp. 61n2, 66.

[3] See, for example, Thomas Nagel, "Rawls on Justice," *The Philosophical Review*, LXXXII, 2 (April 1973): 220–34, at pp. 229–30; John C. Harsanyi, "Can the Maximin Principle Serve as a Basis for Morality? A Critique of John Rawls's Theory," *The American Political Science Review*, LXIX, 2 (June 1975): 594–606, at pp. 598–600; and Derek Parfit, *On What Matters*, vol. 1, ed. Samuel Scheffler (New York: Oxford University Press, 2011), pp. 350–51.

[4] John C. Harsanyi, "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility," *Journal of Political Economy*, LXIII, 4 (August 1955): 309–21, at p. 316.

[5] JF, p. 106. A similar distinction was put forward in Frank H. Knight, *Risk, Uncertainty, and Profit* (Boston: Houghton Mifflin, 1921), pp. 19–20.

[6] Abraham Wald, *Statistical Decision Functions* (New York: Wiley, 1950), p. 18.

Rawls rejects the implausible view that the Maximin Rule for Choice under Uncertainty would be rational for choice under risk.[7] And he does not accept this principle as a *general* principle for rational decisions in all cases of uncertainty.[8] Crucially, Rawls does not require that the parties rely on the Maximin Rule for Choice under Uncertainty in the Original Position.[9]

1. The *Ex-Post* Difference Principle

The Difference Principle can be read in at least two very different ways, depending on whether we evaluate social value *ex post*: with information about how risky prospects turn out, or *ex ante*: without such information, relying instead on expectations.[10] While Rawls favors an [**p. 590**] *ex-ante* version of the Difference Principle, we shall begin with the *ex-post* approach. On this approach, the Difference Principle amounts to the following:

> *The Ex-Post Difference Principle*
> Let the social value of a final outcome be equal to the minimum well-being of any person in the outcome. And let the social value of a prospect be equal to the expected social value of its final outcome. Choose a prospect with a maximal social value among all alternative prospects.[11]

---

[7] John Rawls, "Some Reasons for the Maximin Criterion," *American Economic Review*, LXIV, 2 (May 1974): 141–46, at p. 142; and JF, p. 97n19.

[8] TJ (1971), p. 153, (1999), p. 133; and JF, pp. xxvii, 97n19.

[9] JF, p. 99.

[10] The *ex-ante*/*ex-post* distinction is due to Gunnar Myrdal, *Monetary Equilibrium* (London: Hodge, 1939), p. 47.

[11] This version of the Difference Principle mirrors the maximin structure of *the Maximin Equity Criterion*, according to which a first distribution is socially at least as good as a second distribution if and only if the worst off in the first distribution are at least as well off as the worst off in the second distribution. *The Leximin Equity Criterion*, first suggested by Sen (*Collective Choice and Social Welfare*, *op. cit.*, p. 138n12), is just like the Maximin Equity Criterion except in cases where the worst off in the distributions are equally well off. In those cases, the Leximin Equity Criterion compares the distributions with one of the worst off removed in each distribution. Then, if the worst off among those who remain are better off in one of the distributions, that distribution is socially better than the other. If not, repeat this procedure again until one distribution comes out as socially better or all people who remain are equally well off, in which case the distributions are socially equally good. In TJ (1971), pp. 82–83, Rawls accepts the Leximin Equity Criterion, but—in TJ (1999), p. 72—he claims that the differences between the maximin and leximin criteria do not matter in practice. Likewise, these differences will not matter for the argument of this paper. One noteworthy difference between these criteria, however, is that the Leximin Equity Criterion evaluates final outcomes in terms of a lexical ordering, and lexical orderings cannot be represented by real-valued functions. Since the standard expected-utility approach to calculating expectations

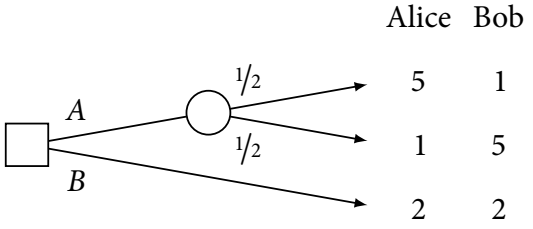
Note that, in Rawls's theory, the Difference Principle is subordinate to the Principle of Justice (demanding equal basic liberties), the Principle of Fair Equality of Opportunity (demanding public offices and social positions to be open to all), and the Just Savings Principle (demanding sufficient savings for the future).[12] For the purposes of our discussion, we can ignore these complications. In the cases we shall discuss, assume that all members of society have equal basic liberties and fair equality of opportunity and that just savings have been made, so that the Difference Principle will apply.

Furthermore, in Rawls's version of the Difference Principle, the relevant comparisons for identifying the least advantaged are made in terms of primary goods.[13] For the sake of brevity, I shall make these **[p. 591]** comparisons in terms of well-being. This is not a substantial change: the well-being levels can represent indexes of primary goods.[14]

Finally, the Difference Principle is only supposed to be applied to the choice of the basic structure of society. The basic structure of society is the way in which fundamental rights and duties are distributed by major social institutions and the way these institutions determine the distribution of advantages from social cooperation.[15] So, in the cases we shall discuss, the choices should be understood as choices determining this basic structure.

To see how the *Ex-Post* Difference Principle works, consider

Case 1



Here, the box represents an initial choice node, where we have a choice between two basic structures of society, *A* (chosen by going up in the choice node) and *B* (chosen by going down). If *A* is chosen, we reach a chance node, represented by the circle, where there is a one-in-two probability that chance goes up, which would give Alice a well-being of 5

requires an evaluation of final outcomes represented by a real-valued function, there is no straightforward way to define an *ex-post* version of the Leximin Equity Criterion.

[12] TJ (1971), pp. 302–03, (1999), pp. 266–67; and JF, p. 61.

[13] TJ (1971), pp. 90–95, (1999), pp. 78–81.

[14] Although I will assume for simplicity that expectations are calculated according to expected utility theory, my argument does not rely on this assumption. And my argument is not vulnerable to the possibility of diminishing marginal value. See appendix.

[15] TJ (1971), p. 7, (1999), p. 6.

and Bob a well-being of 1, and a one-in-two probability that chance goes down, which would give Alice a well-being of 1 and Bob a well-being of 5. If *B* is chosen, everyone is certain to get a well-being of 2. We suppose that the probabilities in the chance node have an objective basis. And, while we shall treat Alice and Bob as two individuals, they could also be thought of as representatives from two complementary halves of society.

In this case, the *Ex-Post* Difference Principle prescribes *B*, because, if we choose *A*, the expected minimum well-being is 1 and, if we choose *B*, the minimum well-being is 2, which is better. Yet choosing *B* gives everyone an expected well-being of 2, whereas choosing *A* gives everyone an expected well-being of 3. Hence the *Ex-Post* Difference Principle violates the following dominance principle: [**p. 592**]

> *The Weak Ex-Ante Pareto Principle*
> If each person has a higher expected well-being in prospect *x* than in prospect *y*, then *y* is not chosen over *x*.[16]

In cases where a principle violates the Weak *Ex-Ante* Pareto Principle, the parties know that, once the Veil of Ignorance has been lifted, they would

[16] Marc Fleurbaey and Alex Voorhoeve put forward an argument against the Weak *Ex-Ante* Pareto Principle in "Decide as You Would with Full Information! An Argument against *Ex Ante* Pareto," in Nir Eyal et al., eds., *Inequalities in Health: Concepts, Measures, and Ethics* (New York: Oxford University Press, 2014), pp. 113–28, at p. 114. Much the same argument can be found in Wlodek Rabinowicz, "Prioritarianism for Prospects," *Utilitas*, XIV, 1 (March 2002): 2–21, at p. 11. The argument is that, combined with some egalitarian principles, the Weak *Ex-Ante* Pareto Principle can violate

> *The Principle of Full Information*
> When one lacks information, but can infer that there is a particular alternative one would invariably regard as best if one had full information, then one should choose this alternative.

Note, however, that the Weak *Ex-Ante* Pareto Principle would only violate this requirement in combination with certain principles; it would not do so in combination with some others. Combined with the Principle of Average Utility, for example, the Weak *Ex-Ante* Pareto Principle would not violate the Principle of Full Information. So it is not clear that the Weak *Ex-Ante* Pareto Principle would be to blame if the Principle of Full Information were violated. In combination with the Difference Principle, the Principle of Full Information prescribes *B* in Case 1, contrary to the Weak *Ex-Ante* Pareto Principle. But, as Fleurbaey and Voorhoeve ("Decide as You Would with Full Information!," *op. cit.*, p. 117) point out, the Principle of Full Information does not seem plausible given the role of the Veil of Ignorance. Fleurbaey and Voorhoeve's (*ibid.*, p. 116) argument relies on an assumption about the agent being "an egalitarian who rightly cares both about reducing outcome inequality and about increasing individuals' well-being." In the Original Position, however, the parties are supposed to try to advance their individual interests; they are not supposed to be concerned about egalitarianism. The principles of justice are what the parties, trying to advance their own individual interests, would agree to; these principles are not what the parties are supposed to be concerned with primarily—see TJ (1971), pp. 118–19, (1999), pp. 102–03. Hence, for the parties in the Original Position, Fleurbaey and Voorhoeve's objection to the Weak *Ex-Ante* Pareto Principle could not get off the ground.

(no matter who they turn out to be) prefer some alternative principle which would give everyone a better expectation.[17]

In Case 1, Alice and Bob would each have a higher expected well-being if, instead, a principle that prescribes *A* were followed. Since it would not be in anyone's interests were the *Ex-Post* Difference Principle followed in Case 1, the parties in the Original Position know that, in that case, were the *Ex-Post* Difference Principle followed, it **[p. 593]** would not be in *their* interests. By this argument, the parties in the Original Position can figure out (without assigning probabilities to their turning out to be any particular person in society) that it would not be in their interests to agree to the *Ex-Post* Difference Principle.

The above argument also applies to a stricter maximin variant of the Difference Principle. Consider

> *The Strict Maximin Difference Principle*
> Let the social value of a final outcome be equal to the minimum well-being of any person in the outcome. And let the social value of a prospect be equal to the minimum social value of any possible final outcome of the prospect. Choose a prospect with a maximal social value among all alternative prospects.

This version of the Difference Principle yields the same result as the *Ex-Post* Difference Principle in Case 1. To see this, note that, if we choose *A*, the minimum possible well-being level is 1 but, if we choose *B*, the minimum possible well-being level is 2, which is better. So, like the *Ex-Post* Difference Principle, the Strict Maximin Difference Principle prescribes *B* in Case 1. Hence it is vulnerable to the same objection as the *Ex-Post* Difference Principle.

It may be objected that the Weak *Ex-Ante* Pareto Principle is only plausible if the parties in the Original Position are risk neutral whereas Rawls seems to assume that the parties are risk averse. But this is neither Rawls's view nor a plausible view. While Rawls's early work might suggest this reading, he later clarified that his argument makes no assumptions about the parties being risk averse, which he agrees would make his argument very weak.[18] On the contrary, Rawls rules out that the parties have any special, non-standard attitudes to risk.[19] He assumes that the

[17] An example of a principle that would have given everyone a higher expected well-being in this case is the Principle of Average Utility. Note, however, that my argument does not rely on this principle. We only need to show that the parties would favour some other principle over the Difference Principle. The parties would, for example, compare the *Ex-Post* Difference Principle unfavorably with a principle that is equivalent except that it prescribes *A* in Case 1.

[18] John Rawls, "Reply to Alexander and Musgrave," *The Quarterly Journal of Economics*, LXXXVIII, 4 (November 1974): 633–55, at pp. 649–50; and JF, pp. xvii, 99, 110.

[19] TJ (1999), p. 148. Compare with TJ (1971), p. 172.

parties are rational in the standard economic sense, being risk neutral.[20] While we shall assume that the parties are risk neutral, my argument does not need this assumption; it only needs to rule out that the parties may have an extreme aversion to risk (see appendix).

One could, for example, resist my argument if one held that the Maximin Rule for Choice under Uncertainty is a principle of rationality for acting under both risk and uncertainty, because it would then be in Alice's and Bob's interests that *B* is chosen in Case 1. But this is not a plausible view. Using the Maximin Rule for Choice under Uncertainty as a way to deal with risk and uncertainty forces us to **[p. 594]** mitigate the worst possible outcome however unlikely, regardless of the likely costs.[21] And, as Rawls points out, this seems irrational.[22]

It may next be objected that Rawls seems to argue that the parties must ignore all probabilities in the Original Position. This would favor the Strict Maximin Difference Principle, since it does not rely on any probabilities. But this is a misreading of Rawls and a misunderstanding of the Veil of Ignorance. Rawls merely objects to the idea that the parties may assign an equal probability to their turning out to be anyone by applying

> *The Principle of Insufficient Reason*
> If there is insufficient reason to regard either of two alternative possibilities as more probable than the other, then they may be regarded as equally probable.[23]

If the parties applied this principle and assigned an equal probability to being anyone, they would maximize their expected well-being by agreeing to

---

[20] JF, p. 87.

[21] See the examples in Harsanyi, "Can the Maximin Principle Serve as a Basis for Morality?," *op. cit.*, pp. 595–96.

[22] Rawls, "Some Reasons for the Maximin Criterion," *op. cit.*, p. 142; and JF, p. 97n19.

[23] The principle should be restricted to a privileged partitioning of possibilities in order to avoid counter-examples of the kind in John Maynard Keynes, *A Treatise on Probability* (New York: Macmillan, 1929), pp. 42–44. While the exact form of this restriction is unclear, the main rival principle for choice under uncertainty faces much the same problem: The Leximin Rule for Choice under Uncertainty is likewise sensitive to the partitioning of possibilities into states of nature (and the Maximin Rule for Choice under Uncertainty ignores improvements in any possible outcome except the worst); see Salvador Barbarà and Matthew Jackson, "Maximin, Leximin, and the Protective Criterion: Characterizations and Comparisons," *Journal of Economic Theory*, XLVI, 1 (October 1988): 34–44, at p. 40. Barbarà and Jackson's own proposal, the Protective Criterion, violates the transitivity of 'equally good as'; *ibid.*, p. 41.

> *The Principle of Average Utility*
> Choose a prospect with a maximal average expected well-being among all alternative prospects.[24]

In his discussion of this argument for the Principle of Average Utility, Rawls does not object to the parties' relying on probabilities that are based on particular facts about society; he merely objects to the use of the Principle of Insufficient Reason. Rawls writes:

> 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.[25]

[p. 595] Put in terms of his risk/uncertainty distinction, Rawls rules out the assigning or estimating of probabilities in cases of uncertainty (where there is no objective basis for estimating probabilities) but not in cases of risk (where there is an objective basis for estimating probabilities).[26] The motivation for this requirement is that the parties in the Original Position should not try to estimate the very knowledge the Veil of Ignorance is supposed to hide.[27] That is why Rawls objects to the parties' using the Principle of Insufficient Reason to estimate the probability of their turning out to be any particular member of society. Rawls's requirement does not demand that the parties ignore probabilities about risky prospects with an objective basis which society and its individuals might face after the Veil of Ignorance is lifted. Those risks are part of what a principle of distributive justice should cover. Unlike probabilities for turning out to be any particular person, which are hidden to ensure impartiality, there are no grounds for ruling out probabilities based on particular facts about risks in society.[28] **[p. 596]**

---

[24] Harsanyi, "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility," *op. cit.*, p. 316.

[25] TJ (1999), p. 149. TJ (1971), p. 173, has a somewhat different wording. See also TJ (1971), p. 168, (1999), pp. 145–46.

[26] JF, p. 106. Rawls describes the interpretation of rationality in the Original Position as "taking effective means to ends with unified expectations and objective interpretation of probability"; TJ (1971), p. 146, (1999), p. 127.

[27] TJ (1971), p. 171, (1999), p. 147.

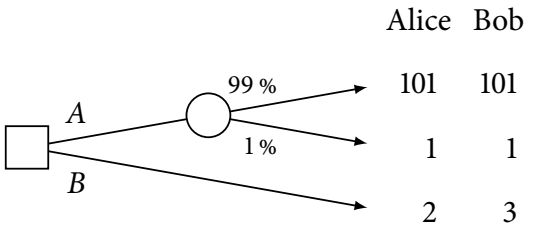
[28] There is one perplexing passage that might seem to conflict with this reading: Rawls states—in TJ (1971), p. 155—that

> the veil of ignorance excludes all but the vaguest knowledge of likelihoods. The parties have no basis for determining the probable nature of their society, or their place in it. Thus they have strong reasons for being wary of probability calculations if any other course is open to them.

Rawls's revision of this passage—in TJ (1999), p. 134—is even stronger, stating that

Still, one may be unconvinced and object that, even though Rawls does not hold this view, he *should* reject any probabilities in the Original Position and favor the Strict Maximin Difference Principle. But, in addition to the above reasons why Rawls rejects this principle, there is a further reason why this principle is an implausible account of justice: The Strict Maximin Difference Principle yields excessively anti-egalitarian results when risks are taken jointly. Consider

Case 2



Here, in the choice node, represented by the box, we have a choice between two basic structures of society, *A* and *B*. If we choose *A*, we would reach a chance node, represented by the circle, where the probability that chance goes up is 99 percent—giving everyone a well-being of 101—and the probability that chance goes down is 1 percent—giving everyone a well-being of 1. If we choose *B*, Alice would get a well-being of 2 whereas Bob would

---

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

(See also JF, p. 98.) In this passage, it might seem that Rawls rules out all deliberation based on probabilities and risks in the Original Position. The problem is that, if one were to rule out all such deliberations, the parties would not be in a position to assess the principles of distributive justice in so far as they cover the distribution of risks in society. For example, if the parties had no knowledge of probabilities, they could not assess whether an *ex-ante* approach would be preferable to an *ex-post* approach. And then, crucially for Rawls, the parties could not be in a position to agree to the *Ex-Ante* Difference Principle, because there would be no way for them to assess what is to the greatest *expected* benefit of the least advantaged members of society and see the advantages of that principle over the *Ex-Post* Difference Principle or even the Strict Maximin Difference Principle. The most plausible reading of the passage is that Rawls stresses that the parties must deliberate under complete uncertainty about the nature of their *actual* society and their place in it; so they may not assign or estimate probabilities to what society and their place in it are actually like. But, since the parties are to agree to *general* principles of distributive justice, they need to (and may) consider the possible risks in all hypothetical choices covered by these principles for all hypothetical societies that they could (as far as they know) be part of. Being able to reason about these hypothetical probabilities with a hypothetical objective basis is consistent with the parties having 'no basis for determining the probable nature of their society', since they deliberate under uncertainty regarding which one of these hypothetical societies they actually live in. So the last sentence of the revised passage should probably be read as "Thus they have no basis for probability calculations [about the society they actually live in]."

get a well-being of 3. Like before, we assume that these probabilities have an objective basis. In this case, the Strict Maximin Difference Principle prescribes *B*. Yet *B* has an unequal outcome, whereas the outcome of *A* is perfectly equal both *ex ante* and *ex post*. The risk we would take if we chose *A* would be shared by everyone equally and be to everyone's expected benefit: *A* gives everyone an expected well-being of 100, whereas *B* gives Alice and Bob an expected well-being of 2 and 3 respectively. To favor the unequal prospect of *B* in Case 2 on the grounds of justice is to confuse justice with risk aversion. Thankfully, Rawls does not hold this view. [**p. 597**]

2. The *Ex-Ante* Difference Principle

As we have seen, the parties in the Original Position would reject the *Ex-Post* Difference Principle. And, as mentioned, Rawls also rejects that principle. Rawls maintains that social and economic inequalities must be

> to the greatest expected benefit of the least advantaged members of society (the maximin equity criterion)[29]

This suggests

> *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 version of the Difference Principle avoids the problematic implications of the *ex-post* approach in Case 1. The *Ex-Ante* Difference Principle prescribes *A* in Case 1, because *A* maximizes the minimum expected well-being: The minimum expected well-being level if we choose *A* is 3, but, if we choose *B*, it is 2.

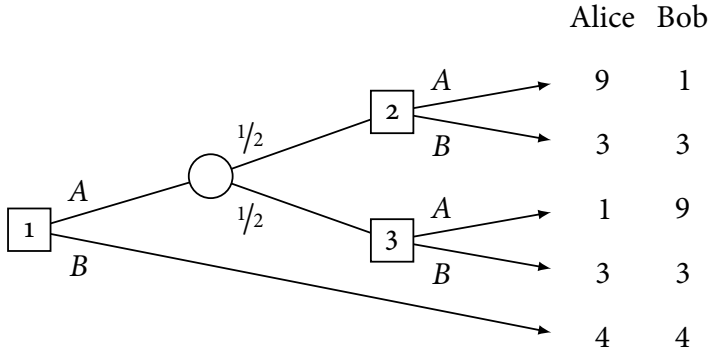
Likewise, the *Ex-Ante* Difference Principle avoids the problematic implications of the Strict Maximin Difference Principle in Case 2. The

[29] Rawls, "Some Reasons for the Maximin Criterion," *op. cit.*, p. 142. See also Rawls's first statement of the Difference Principle, in Rawls, "Distributive Justice," *op. cit.*, p. 66. Rawls's statement in TJ (1971), p. 83, leaves out 'expected', but his revised statement in TJ (1999), p. 72, includes it. Yet—in both TJ (1971), p. 92, and (1999), p. 79—Rawls clearly favors an *ex-ante* approach, stating that the comparisons for the application of the Difference Principle "are made in terms of expectations of primary social goods." In JF, pp. 42–43, Rawls also leaves out 'expected' in the statement of the Difference Principle, but he clarifies (JF, p. 59) that "the inequalities to which the difference principle applies are difference in citizens' (reasonable) expectations of primary goods over a complete life." In John Rawls, *Political Liberalism* (New York: Columbia University Press, 1993), p. 6, (hereinafter referred to as PL) Rawls first states the principle without 'expected' but later (PL, p. 271) with 'expected'. This strongly suggests that the principle should be read with an implicit 'expected' even when Rawls, for some unknown reason, leaves it out.

Ex-Ante Difference Principle prescribes *A* in Case 2, because choosing *A* maximizes the minimum expected well-being: The minimum expected well-being level is 100 if we choose *A*, but, if we choose *B*, it is 2. Hence the *Ex-Ante* Difference Principle is not open to the earlier objections to the *ex-post* approach.

Nevertheless, consider the following sequential case: **[p. 598]**

Case 3



In this case, there are three choice nodes, represented by the numbered boxes. And there is a chance node, represented by the circle, where there is a one-in-two probability that chance goes up and a one-in-two probability that chance goes down. Like before, we assume that these probabilities have an objective basis. Choice node 1 is a first choice between two basic structures *A* and *B*. Choice nodes 2 and 3 are later opportunities to revise the first choice. In each choice node, *A* is chosen by going up and *B* is chosen by going down.

The plan to adopt and stick to *A* in choice node 1 has a minimum expected well-being of 5, since for both Alice and Bob that plan would amount to a fifty-fifty gamble between getting a well-being of 1 or 9 (giving them both an expected well-being of 5). The plan to adopt and stick to *B* in choice node 1 has a minimum expected well-being of 4, since it would give each of Alice and Bob a well-being of 4. So, if we assess these basic structures with the *Ex-Ante* Difference Principle in choice node 1, it seems that we should choose *A*, since it maximizes the minimum expected well-being. Choosing *A* requires that we go up in choice node 1. And, if we were to go up in choice node 1, then, depending on chance, we would face either choice node 2 or choice node 3.

Suppose we face one of choice nodes 2 and 3. These choice nodes also offer a choice between basic structures, as they offer an opportunity to revise the earlier choice between *A* and *B*. So we should consult the *Ex-Ante* Difference Principle again. In choice nodes 2 and 3, *A* has a minimum expected well-being of 1, since it gives one of Alice and Bob a well-being of 9 and the other a well-being of 1. And *B* has a minimum expected well-being of 3, since it gives each of Alice and Bob a well-being of 3. Assessing

these basic structures with the *Ex-Ante* Difference Principle in choice nodes 2 and 3, we should choose *B* rather than *A*, since *B* maximizes the minimum expected well-being.

So, by continuously applying the *Ex-Ante* Difference Principle in Case 3, we would first choose *A* in choice node 1 and then, in one of **[p. 599]** choice nodes 2 and 3, we would revise the basic structure of society to *B*, giving everyone a well-being of 3. This would be wrong: It makes everyone worse off than they would have been if *B* had been chosen in choice node 1, which would have given everyone a well-being of 4.

At this point, it may be objected that the problem here is not the *Ex-Ante* Difference Principle but only this myopic application of that principle—that is, applying it without taking into account what it would prescribe in future choice nodes. Therefore, let us combine the *Ex-Ante* Difference Principle with *backward induction*, which is to first consider what would be chosen in later choice nodes and then take the predicted choices into account when we consider earlier choices. As we have seen, the *Ex-Ante* Difference Principle prescribes *B* in choice nodes 2 and 3. Taking this into account at choice node 1, choosing *A* gives each of Alice and Bob an expected well-being of 3, but choosing *B* gives each of Alice and Bob an expected well-being of 4. So the *Ex-Ante* Difference Principle applied with backward induction prescribes *B* in choice node 1.

Thus, in Case 3, the *Ex-Ante* Difference Principle results in either everyone getting a well-being of 3 (applied myopically) or everyone getting a well-being of 4 (applied with backward induction). Either way, the *Ex-Ante* Difference Principle does worse in Case 3 than a principle that prescribes choosing and sticking to *A*, that is, to follow the plan of choosing *A* in all three choice nodes. Choosing and sticking to *A* gives each of Alice and Bob an expected well-being of 5, since it would amount to a fifty-fifty gamble for each between getting a well-being of 1 or 9. So following the *Ex-Ante* Difference Principle in Case 3 gives everyone an expected well-being of either 3 or 4, but following an alternative principle that prescribes choosing and sticking to *A* gives everyone an expected well-being of 5.[30]

Hence the *Ex-Ante* Difference Principle violates

> *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 expected outcome is *y* is not followed if there is an alternative plan available whose expected outcome is *x*.

[30] An example of a principle that would prescribe choosing and sticking to *A* in Case 3 is the Principle of Average Utility.

This violation illustrates that it would not be in anyone's rational interests that the *Ex-Ante* Difference Principle were followed in Case 3.[31] **[p. 600]** The point of the *Ex-Ante* Difference Principle is to arrange the basic structure of society to the expected benefit of the least advantaged. But, as we have seen in sequential cases, this principle can lower the expectations of the least advantaged. As Rawls writes,

> a principle is ruled out if it would be self-contradictory, or self-defeating, for everyone to act upon it. . ..Principles are to be chosen in view of the consequences of everyone's complying with them.[32]

Accordingly, the parties in the Original Position would not agree to the *Ex-Ante* Difference Principle, since in Case 3 they know that—no matter who they are in society—it would not be in their interest to adopt that principle. By this argument, the parties in the Original Position are led to reject the *Ex-Ante* Difference Principle without being able to assign or estimate probabilities to their turning out to be any particular member of society.

It may be objected that the basic structure of society only needs to be chosen once.[33] And, if so, there would be no need to revise the basic structure at choice nodes 2 and 3. So one could apply the *Ex-Ante* Difference Principle myopically in choice node 1, choose *A*, and then simply keep that structure. The problem with this move is that the justification for *A* in choice node 1 is that *A* is prescribed by the *Ex-Ante* Difference Principle, but this justification no longer applies in choice nodes 2 and 3, since, in those nodes, the *Ex-Ante* Difference Principle prescribes *B*.[34] **[p. 601]**

There is, however, a variation of the *Ex-Ante* Difference Principle which ensures that the minimum expected well-being would be maximized consistently relative to a privileged node (or point in time). This variation focuses, at all times, on the plans that were available in the

[31] At least, it would not be in anyone's long-term lifetime interest, which is what matters according to Rawls, TJ (1971), p. 64, (1999), p. 56; and JF, p. 59. This focus on lifetime well-being is what blocks the sequential argument against the Difference Principle in D. W. Haslett, "Does the Difference Principle Really Favour the Worst Off?," *Mind*, XCIV, 373 (January 1985): 111–15, at pp. 111–12.

[32] TJ (1971), p. 132, (1999), p. 114.

[33] I thank Krister Bykvist for raising this objection. Rawls, however, maintains that the basic structure would need adjustments even in a well-ordered society. Even if the principles of justice remain the same, technology and other circumstances may change, which may change what basic structure is the best implementation of the unchanged principles of justice. See John Rawls, "The Basic Structure as Subject," *American Philosophical Quarterly*, XIV, 2 (April 1977): 159–65, at p. 164; and PL, p. 284.

[34] Note moreover that, although the initial choice of basic structure in Case 3 helps the presentation, it is inessential to the argument. To see this, consider the following variation without the first choice node:

privileged node. Here, a plan that is available in the privileged node is a specification of what to choose in each choice node that can be reached from the privileged node. Consider

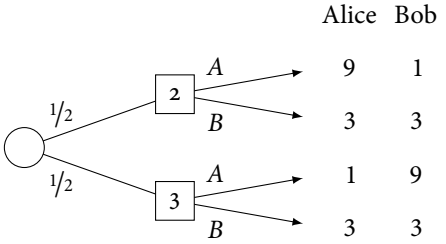
> *The Resolute Ex-Ante Difference Principle*
> Let the social value of a plan be equal to the minimum expected well-being of any person if the plan were followed, with expectations calculated from a certain privileged initial node. Choose a prospect following a plan with maximal social value among the plans that (i) were available in the privileged node and (ii) are still feasible.[35]

The Resolute *Ex-Ante* Difference Principle demands that one follow a plan that maximizes the minimum expected well-being relative to the privileged node. In Case 3, if we let choice node 1 be the privileged node, the alternative plans in that node will be valued by their minimum expected well-being as follows:

- *A in choice node 1; A in choice node 2; A in choice node 3*
  Minimum expected well-being: 5
- *A in choice node 1; A in choice node 2; B in choice node 3*
  Minimum expected well-being: 2
- *A in choice node 1; B in choice node 2; A in choice node 3*
  Minimum expected well-being: 2
- *A in choice node 1; B in choice node 2; B in choice node 3*
  Minimum expected well-being: 3
- *B in choice node 1*
  Minimum expected well-being: 4

---

*Case 3**



In this variation, we have that, calculated from the initial chance node, following the *Ex-Ante* Difference Principle in the choice nodes gives everyone an expected well-being of 3, whereas following a principle that prescribes *A* in these choice nodes (such as the Principle of Average Utility) gives everyone an expected well-being of 5. In this variation, the basic structure of society is only chosen once. Yet the parties can still see that it is not in their individual interests to agree to the *Ex-Ante* Difference Principle.

[35] The resolute approach is based on McClennen's resolute-choice decision theory in Edward F. McClennen, *Rationality and Dynamic Choice: Foundational Explorations* (New York: Cambridge University Press, 1990), p. 13.

So, if choice node 1 is the privileged node, the Resolute *Ex-Ante* Difference Principle prescribes that one follow the first plan of choosing *A* in all three choice nodes. If one follows this plan, one avoids choosing so that everyone gets a worse expected well-being in choice node 1 than they could have had if one had followed an alternative plan. Note, however, that sticking to the first plan involves not benefiting the least advantaged in one of choice nodes 2 and 3. [**p. 602**]

Yet the main problem with the Resolute *Ex-Ante* Difference Principle is its need for a privileged node or time. In choice node 2 (or 3), if that node were the privileged node, the Resolute *Ex-Ante* Difference Principle would prescribe *B*. Yet, as we saw earlier, if choice node 1 were the privileged node, then the principle would prescribe *A* in choice node 2 (or 3). The problem is that no time could plausibly serve as a non-arbitrary privileged time in the Original Position.

One suggestion for a privileged time could be the start or founding of society. But, first, there is typically no exact point in time at which a society is founded, and it seems to some extent arbitrary how societies should be individuated over time. So any specific, exact time for the founding of society would be arbitrary. Second, it seems that the time of the founding of society would only be significant to the parties if they had some reason to think that they entered the Original Position at that time. After the founding of society (in particular for later generations), the parties have no reason to attach any significance to expectations calculated relative to the time of the founding. Their concern, trying to advance their individual interests, would be their potential expectations after the veil is lifted—that is, their expectations relative to the time they entered, or will exit, the Original Position. Third, it is not clear that people who belong to later generations would have any meaningful expectations calculated at the founding of society if it was still uncertain at that time whether they would ever be born, because those who are never born in some potential outcome might lack a well-being level in that outcome. Expectations of well-being require a well-being level for each potential outcome.

Another suggestion is to have a separate Original Position for each new generation, each generation choosing its own separate privileged node for the Resolute *Ex-Ante* Difference Principle. Generations, however, are continuous: there is no non-arbitrary time at which a new generation starts. Moreover, generations overlap; so the Resolute *Ex-Ante* Difference Principle needs to cover distributions between contemporary yet distinct generations. And, with different privileged nodes, we could get incompatible prescriptions. Consider, for example, Case 3, and suppose that Alice and Bob belong to two separate yet overlapping generations and that one generation enters the Original Position at the time of choice

node 1 and the other enters at the time of choice node 2 (or 3). Given that choice node 1 is the privileged node, the Resolute *Ex-Ante* Difference Principle prescribes *A* in choice node 2. But, given that choice node 2 is the privileged node, the principle disallows *A* in choice node 2.

A more general problem is that any time-sensitive manner of picking a privileged node would require time-sensitive information in the Original Position. This conflicts with Rawls's specification that **[p. 603]**

> the original position must be interpreted so that one can at any time adopt its perspective. It must make no difference when one takes up this viewpoint, or who does so: the restrictions must be such that the same principles are always chosen. The veil of ignorance is a key condition in meeting this requirement. It insures not only that the information available is relevant, but that it is at all times the same.[36]

If principles of justice are justified via the Original Position, it seems that the principles that are justified at a time are those principles that would be agreed to in the Original Position if it were (hypothetically) entered at that time.[37] But, if the choice of these principles were based on time-sensitive information, different principles would be chosen (and thus justified) at different times. While the basic structure of society may plausibly need revision from time to time, it is implausible that the underlying principles of justice would change.[38] If the parties knew the time of their entry into the Original Position and picked the privileged point based on that information, their choice would be time sensitive contrary to Rawls's specification. But, if they do not know the time of their entry into the Original Position, there seems to be no non-arbitrary time they could be in a position to pick as the privileged one. Hence, like the other versions of the Difference Principle, the Resolute *Ex-Ante* Difference Principle would not be chosen by the parties in the Original Position.

3. Conclusion

As we have seen, there are several versions of the Difference Principle, and they are open to very different problems. No matter which version we pick,

[36] TJ (1999), p. 120. The wording in TJ (1971), p. 139, is slightly different.

[37] TJ (1971), pp. 19–21, (1999), pp. 17–19; John Rawls, "Justice as Fairness: Political Not Metaphysical," *Philosophy and Public Affairs*, XIV, 3 (Summer 1985): 223–51, at pp. 237–39; John Rawls, "The Basic Structure as Subject," in Alvin I. Goldman and Jaegwon Kim, eds., *Values and Morals: Essays in Honor of William Frankena, Charles Stevenson, and Richard Brandt* (Boston: Reidel, 1978), pp. 47–71, at p. 59; and PL, pp. 274–75.

[38] Rawls claims that "first principles must be capable of serving as a public charter of a well-ordered society in perpetuity"; TJ (1971), p. 131, (1999), pp. 113–14. Having separate versions of the Resolute *Ex-Ante* Difference Principle for different generations seems to violate this requirement.

however, we have seen that we would face one of two problems. Either it would not be in the interests of the parties in the Original Position to adopt the Difference Principle in at least one of Cases 1, 2, and 3, or the principle would need to refer to a privileged time, which would exclude it from the discussions behind the Veil of Ignorance. Hence the parties in the Original Position would not agree to the Difference Principle. [**p. 604**]

Appendix

For simplicity, I have assumed that the value of expectations are calculated according to expected utility theory. This may have raised some worries about risk aversion and diminishing marginal value of well-being if well-being levels represent indexes of primary goods.

But all that is needed for my discussion of Cases 1 and 3 is that there are three levels *a*, *b*, and *c* such that *a* is better than *b*, *b* is better than *c*, and a gamble with a one-in-two probability of *a* and a one-in-two probability of *c* is a better expectation than *b* with certainty. This assumption does not conflict with diminishing marginal values. To see this, consider monetary expectations. Even though $2,000 is not twice as good as $1,000, it is still very plausible that a gamble with a one-in-two probability of $2,000 and a one-in-two probability of $1,000 is a better expectation than $1,001 with certainty. This is plausible because the difference in value between getting $2,000 and getting $1,001 (that is, the potential gain from the gamble) is still much larger than the difference in value between getting $1,001 and getting $1,000 (that is, the equally likely potential loss).

So we only need three levels *a*, *b*, and *c* such that *a* is better than *b*, *b* is better than *c*, and the difference in value between getting *a* and getting *b* is much larger than the difference in value between getting *b* and getting *c*. This requirement can be met even if *a* is only a little bit better than *c*, because we can pick a level *b* such that *b* is only better than *c* by an arbitrarily small amount. Then, in Case 1, we could replace level 5 with *a*, level 2 with *b*, and level 1 with *c*. And, in Case 3, we could replace level 9 with *a*, level 4 with *b*, level 1 with *c*, and level 3 with any level that is worse than *b* but better than *c*. Given a revision of this form, my arguments should be compatible with any non-extreme form of risk aversion.

Likewise, all that is needed for my discussion of Case 2 is that there are three levels *a*, *b*, and *c* such that *a* is better than *b*, *b* is better than *c*, and a gamble with a 99 percent probability of *a* and a 1 percent probability of *c* is a better expectation than *b* with certainty. The only difference to the assumption for the other cases is that the better outcome in the gamble is more probable. Hence the plausibility of this assumption follows by the same kind of argument as before. So, in Case 2, we could replace level 101

with *a*, level 3 with *b*, level 1 with *c*, and level 2 with any level that is worse than *b* but better than *c*.