Rank-Weighted Utilitarianism and the Veil of Ignorance[*]

Jacob M. Nebel

Forthcoming in *Ethics*

Abstract

Lara Buchak argues for a version of rank-weighted utilitarianism that assigns greater weight to the interests of the worse off. She argues that our distributive principles should be derived from the preferences of rational individuals behind a veil of ignorance, who ought to be risk averse. I argue that Buchak's appeal to the veil of ignorance leads to a particular way of extending rank-weighted utilitarianism to the evaluation of uncertain prospects. This method recommends choices that violate the unanimous preferences of rational individuals and choices that guarantee worse distributions. These results, I suggest, undermine Buchak's argument for rank-weighted utilitarianism.

1 Introduction

Consider two distributions in which there are two people: Ann and Bob. In distribution *A*, Ann and Bob are equally well off. In *B*, Ann is better off than Bob. Both contain the same total well-being. These distributions are shown in table 1. The number in each cell represents the welfare of the corresponding person (column) in each distribution (row).[1]

According to utilitarianism, these distributions are equally good. Harsanyi (1953) argues for this answer by appealing to the preferences of a rational "impartial observer" behind a veil of ignorance. The observer is impartial because they do not know their own identity. In Harsanyi's model, they assign an equal probability to being each person and seek to maximize their expected utility (EU)—i.e., the probability-weighted average of their welfare levels in each outcome.

---

[1]I use "welfare," "well-being," and "utility" interchangeably to refer to how good an outcome is for a person.

Table 1: Equal vs. Unequal

	Ann	Bob
Distribution *A*	2	2
Distribution *B*	4	0

The impartial observer faces a choice between the prospects in table 2. They assign an equal probability to two states, displayed in the columns: in state 1, they are Ann; in state 2, they are Bob. Each prospect (row) specifies what the impartial observer's welfare would be under each state (column). The third column displays their expected utility.

Table 2: The Veil of Ignorance

	State 1 (0.5)	State 2 (0.5)	EU
Prospect *A*	2	2	2
Prospect *B*	4	0	2

According to Harsanyi, the impartial observer ought to be indifferent between these two prospects, because they have the same EU. Thus, distributions *A* and *B* must be equally good, because an impartial observer ought to be indifferent between their corresponding prospects.

Buchak (2017) agrees with Harsanyi that we should choose whatever distribution would be preferred by a rational agent who assigns an equal probability to being each person. But she departs from Harsanyi in one crucial respect. She denies that the impartial observer should maximize their EU. Buchak argues that it is rational for individuals to be risk-averse in a way that is incompatible with EU theory. The impartial observer, in particular, *ought* to be risk-averse. This leads Buchak's impartial observer to give greater weight to possibilities in which they are a worse-off person than to possibilities in which they are a better-off person. And this leads Buchak to a distinctive theory of distribution, according to which the interests of the relatively worse off matter more than those of the relatively better off.

Buchak's preferred theory of rational choice is what she calls *risk-weighted expected utility* (REU) theory. REU theory weights outcomes not by their raw probabilities, but by their *risk-weighted* probabilities, where the risk-weighted probability of an outcome depends on its position in the prospect's rank-ordering of outcomes.[2] A risk-avoidant agent assigns

---

[2]This formulation oversimplifies somewhat. Buchak's risk function is applied to *cumulative* probabilities

greater weight to worse outcomes. (EU theory is the special case of REU theory where every outcome receives the same weight. But I will use "REU theory" to refer only to versions that depart from EU theory, by holding that it can be rational to assign different weights to different outcomes.)

According to Buchak, we should ascribe to the impartial observer the most risk-avoidant attitude within reason. Given any risk-avoidant weighting of probabilities, the impartial observer would rationally prefer prospect *A* to *B*. So, according to Buchak, distribution *A* is better than *B*.

Whereas the distributive analogue of EU theory is utilitarianism, the distributive analogue of REU theory is *rank-weighted* utilitarianism (RWU). RWU evaluates distributions according to a weighted average of individual utilities. Each person's utility is multiplied by a weight assigned to their position in the distribution's rank-ordering of utilities. (Utilitarianism is the special case of RWU in which all weights are equal, but I will use "RWU" to refer only to those versions that depart from utilitarianism, by assigning different weights to at least some different people.)

If the impartial observer is an REU maximizer, then they prefer distributions with greater rank-weighted average utility. And if, as Buchak argues, they are risk averse, then they assign greater weight to worse-off positions in a distribution. This characterizes a particular class of rank-weighted utilitarian principles: the *generalized Gini family* (Weymark 1981).

For concreteness, I will work with a particular member of this family: the standard Gini social welfare function, whose weights are the first *n* odd numbers (where *n* is the number of people). So, the welfare of the best-off person gets weighted by one, the welfare of the second-best-off person gets weighted by three, and so on. We normalize these weights by dividing them by their sum (n^2—i.e., the sum of the first *n* odd numbers). So, in a population of two people, the normalized weight applied to the worst-off person is 0.75; the normalized weight applied to the better-off person is 0.25.[3] I assume an analogous risk attitude for the impartial observer: the weights are the first *n* odd numbers (where *n* is the number of states), and we normalize these weighted probabilities to sum to one. My arguments will not depend on this particular choice of weights, but it will help to illustrate the arguments with particular examples.

Now suppose that we—in the actual world, who know our identities—must choose

---

and then weights *differences* in utility. But this will not affect the arguments.

[3] $3/(3 + 1) = 0.75$; $1/(3 + 1) = 0.25$.

between two policies, and that we are uncertain about their distributive outcomes. How should we choose? Buchak herself does not discuss this question. But we might have reason to hope that her unified approach to rational choice and distribution would lead to, or at least be compatible with, an attractive theory of social choice under uncertainty.

In particular, we might reasonably suspect that REU theory would allow RWU to avoid the dilemma posed by Harsanyi's (1955) other argument for utilitarianism: his aggregation theorem. Harsanyi's aggregation theorem runs as follows. Consider a society of EU maximizers. Suppose that social preferences also satisfy EU theory, in the sense that society's preferences between policies can be represented as maximizing the expected value of those policies' outcomes (leaving open how the value of an outcome is determined). And suppose that social preferences respect the unanimous preferences of individuals, in the sense that society prefers any policy that is preferred by everyone. Then, Harsanyi argues, social preferences must be utilitarian.

Here is a simple illustration of the problem, applied to RWU. Consider the social prospects depicted in table 3. Prospect A guarantees that Ann and Bob are equally well off $(3, 3)$. B has an equal probability of making Ann better off than Bob $(8, 0)$ or of making Bob better off than Ann $(0, 8)$, with greater aggregate well-being either way. The "RWU" columns next to each outcome depict the outcomes' rank-weighted average utility. The "EU" columns depict each person's EU in these prospects.

Table 3: RWU with EU Theory

| | State 1 (0.5) | | | State 2 (0.5) | | | EU | |
	Ann	Bob	RWU	Ann	Bob	RWU	Ann	Bob
Prospect A	3	3	3	3	3	3	3	3
Prospect B	8	0	2	0	8	2	4	4

In table 3, prospect A guarantees a better distribution, according to RWU (using the Gini weighting), than prospect B.[4] But prospect B offers both Ann and Bob greater expected well-being. Given EU theory, RWU generates a conflict between respect for individual preferences, which favors B, and social rationality, which favors A.

In response to the argument from Harsanyi's aggregation theorem, most opponents of utilitarianism in distributive ethics tend to retain EU theory for individuals. Some respond by rejecting EU theory for social preferences (Diamond 1967). This response is asso-

---

[4]B guarantees a rank-weighted average of 2 = 1/4(8) + 3/4(0). A guarantees 3 = 1/4(3) + 3/4(3).

ciated with the "ex ante" school of social choice, which takes social welfare to be a function of individuals' prospects, as opposed to their outcomes. (It is called "ex ante" because it can be regarded as applying the social welfare function to the distribution of prospects *before* the resolution of uncertainty.) Others respond by denying that social preferences must promote individuals' expected interests (Adler 2011; Fleurbaey and Voorhoeve 2013). This response is associated with the "ex post" school of social choice, which applies the social welfare function to final outcomes and evaluates social prospects in terms of the probabilities and values of their possible outcomes. (It is called "ex post" because it can be regarded as applying the social welfare function to final outcomes *after* the resolution of uncertainty.) For a utilitarian, the ex ante and ex post approaches can never diverge, because the average of expected utilities is equal to the expectation of average utilities (at least, in fixed-population cases).

Buchak's work suggests a different response: rejecting EU theory, and replacing it with REU theory, for both individual and social preferences. This strategy might seem especially promising for the following reason. We might think that the combination of RWU and EU theory generates the conflict in table 3 because RWU evaluates distributions in a way that is fundamentally disanalogous to the way that EU theory evaluates individual prospects—by simply weighting utilities by their probabilities. The utilitarian, by contrast, seems able to maximize expected value while respecting the unanimous preferences of EU-maximizing individuals because utilitarianism is the straightforward distributive analogue of EU theory. Since RWU is the straightforward distributive analogue of REU theory, we might expect the structural harmony between REU theory and RWU to allow the rank-weighted utilitarian to maximize risk-weighted expected value while respecting the unanimous preferences of REU-maximizing individuals.

Indeed, the conflict in table 3 dissolves if EU theory is replaced by REU theory. Suppose that Ann and Bob have the same risk attitude—the Gini weighting—and that social preferences satisfy REU theory with this same risk attitude as well. On these assumptions, prospect *B* offers both Ann and Bob a risk-weighted expected utility of 2—which is the same as the rank-weighted average utility of *B*'s outcomes.[5] So both Ann and Bob would rationally prefer *A*. Given RWU and REU theory, there is no need to choose between respect for individual preferences and social rationality in table 3. By contrast, given *utilitarianism* and REU theory, we get the same conflict that arose from the combination of RWU and EU theory in table 3. For, according to utilitarianism, *B* guarantees a better outcome than *A*. So, given REU theory, it is the utilitarian who must choose between respect for individual

---

[5]The calculation is the same as in footnote 4.

preferences and social rationality—at least, in this particular case.

Unfortunately, however, this harmony between REU theory and RWU breaks down in other cases. I raise a variant of the problem posed by Harsanyi's aggregation theorem, replacing EU with REU maximization. It turns out that, even when all individuals are REU maximizers with the same attitude towards risk, society cannot always maximize risk-weighted expected value while respecting the unanimous preferences of individuals.[6] I illustrate this by considering three ways of applying RWU under uncertainty. The first, considered in section 2, is an ex post approach, which applies RWU to the distribution of well-being in each of a policy's possible outcomes, and then takes the risk-weighted expectation of the resulting values. This approach violates the unanimous preferences of rational individuals. The second, considered in section 3, is an ex ante approach, which applies RWU to each policy's distribution of risk-weighted expected utilities for each individual, rather than their utilities in each outcome. This approach recommends choices that guarantee worse distributions. I argue, however, that Buchak's appeal to the veil of ignorance leads to a third approach, considered in section 4, which violates the unanimous preferences of rational individuals *and* recommends choices that guarantee worse distributions. This result, I suggest, undermines Buchak's argument for RWU. Section 5 generalizes these results and explains, at a deeper level, why they arise.

2   Ex Post RWU

Perhaps the most obvious way to apply RWU under uncertainty is to combine it with REU theory. According to

Ex Post RWU: We ought to choose one prospect rather than another just in case the one offers a greater risk-weighted expectation of rank-weighted utilitarian value.

On this principle, we first use RWU to assign a value to each possible outcome; we then choose the policy that offers the highest risk-weighted expected value. This principle seems natural given the foundational role of REU maximization in Buchak's argument for RWU. If REU maximization is the correct theory of individual rational choice under uncertainty,

---

[6]Blessenohl (2020) demonstrates an analogous result involving REU maximizers with *different* risk attitudes, with considerably weaker constraints on social preferences than REU and RWU. I show in section 5.2 that this impossibility does not require different risk attitudes. An argument similar to the one in section 5.2 is given by Kowalczyk (2019). For more general results along these lines, Mongin and Pivato (2015) and McCarthy, Mikkola, and Thomas (2020).

then we might reasonably expect it to be the correct theory of moral or social choice under uncertainty as well.

Ex post RWU requires a risk attitude for social choice. We can reasonably assume that, if everyone has the same risk attitude, and if that attitude is reasonable, then society should use that risk attitude. I will suppose, for simplicity, that everyone has the most risk-avoidant attitude within reason—which I supposed above to be given by the Gini weighting—so that society ought to use this risk attitude as well.

Ex post RWU faces the following problem. Consider the options in table 4. There are two individuals, Cat and Dan, and two equiprobable states. If state 1 obtains, then prospect C would make Cat and Dan equally well off $(20, 20)$, and D would make Cat somewhat better off than Dan $(20, 15)$. If state 2 obtains, then C would make Cat much better off than Dan $(50, 10)$, and D would make both Cat and Dan even better off $(60, 20)$. Suppose that both Cat and Dan have the most risk-averse attitude within reason—i.e., the Gini weighting.

Table 4: Ex post RWU violates ex ante Pareto

| | State 1 (0.5) | | | State 2 (0.5) | | | REU | | RE-RWU |
	Cat	Dan	RWU	Cat	Dan	RWU	Cat	Dan	RWU
Prospect C	20	20	20	50	10	20	27.5	12.5	20
Prospect D	20	15	16.25	60	20	30	30	16.25	19.6875

Given our choice of risk weighting, ex post RWU implies that we ought to choose C rather than D. This evaluation is depicted in the "RE-RWU" column, which shows the risk-weighted expectation of the rank-weighted utilitarian values under the "RWU" columns next to each outcome.[7]

Intuitively, however, I take it that we should choose D. Although D introduces inequality when there would otherwise be none in C, this is outweighed by the more significant benefits to both Cat and Dan in state 2. This judgment also follows from

Ex Ante Pareto: If, relative to a shared, reasonable risk attitude and probability distribution, everyone would rationally prefer prospect X to prospect Y, then we ought to choose X rather than Y.

---

[7]C yields a rank-weighted utilitarian value of $0.25(20) + 0.75(20) = 20$ in state 1, $0.25(50) + 0.75(10) = 20$ in state 2, so a risk-weighted expected value of $0.25(20) + 0.75(20) = 20$. D yields $0.25(20) + 0.75(15) = 16.25$ in state 1, $0.25(60) + 0.75(20) = 30$ in state 2, so $0.25(30) + 0.75(16.25) = 19.6875$.

In table 4, *D* offers each person greater REU relative to their shared risk attitude. These risk-weighted expected utilities are displayed in the columns under "REU."[8] So, according to REU theory, everyone would rationally prefer *D* to *C*. So, by ex ante Pareto, we ought to choose *D*—contrary to ex post RWU. It is worth noting, however, that ex ante Pareto favors *D* not only for our choice of risk attitude, but also for any other risk attitude, and indeed according to every plausible decision theory I know of. (This is because, for each person, *D stochastically dominates C*: for any welfare level, the person's probability of being at least as well off as that level is at least as great in *D* as it is in *C*; and, for some welfare level, it is greater.)

Ex post RWU violates ex ante Pareto in table 4 because it is risk averse in a seemingly impersonal way: it is averse to risks that are faced by society but not by any individual. Ex post RWU favors *C* over *D* because *D* is risky—it might generate a rank-weighted average utility of 30, but might instead generate 16.25—whereas *C* guarantees a constant rank-weighted average utility of 20. REU theory at the social level therefore recommends *C*. But this risk is purely social, not individual: it doesn't arise from any individual's having a greater chance of being worse off. Notice, by contrast, that a neutral risk attitude at the social level—i.e., EU theory for social preference—would recommend prospect *D*, because *D* offers a greater expectation of rank-weighted average utility.

How should a rank-weighted utilitarian respond to this problem?

One option would be to reject ex ante Pareto. This option is perhaps most continuous with extant responses to Harsanyi's aggregation theorem from egalitarians and prioritarians. But it seems at odds with Buchak's argument for RWU, which appeals to the veil of ignorance. Behind the veil of ignorance, we aim to maximize our REU but do not know which person we are. If we know that some prospect offers every person lower REU than every other available policy, then we know that it fails to maximize our REU, since we must be one of those people. So we should not prefer it.

To justify her appeal to the veil of ignorance, Buchak claims that we should make choices that are justifiable to each person. She argues that

> if no reasonable person would reject an option on the grounds that it is too risky, then we are justified in choosing that option. Conversely, if a reasonable person could reject it on these grounds, then we are not justified in choosing it. (p. 631)

---

[8]Cat's REU under *C* is $0.75(20) + 0.25(50) = 27.5$. Cat's under *D* is $0.75(20) + 0.25(60) = 30$. Dan's under *C* is $0.75(10) + 0.25(20) = 12.5$. Dan's under *D* is $0.75(15) + 0.25(20) = 16.25$.

No reasonable person would reject prospect *D* on the grounds that it is too risky. So it seems that we should be justified in choosing *D*. And both Cat and Dan could reasonably reject prospect *C*. So it seems that we should not be justified in choosing *C*. Buchak adds that

> the way we justify our policy or distribution to each citizen is this: if you didn't know the things that cloud your moral judgment—namely, which individual you are, what characteristics you have, and what your actual risk attitude happens to be—then you yourself would have chosen this policy or distribution. (p. 635)

These remarks suggest that we should not choose an option that minimizes every person's REU. Such an option cannot be justified to anyone, let alone everyone.

A rejection of ex ante Pareto could be justified by a broadly impersonal distributive theory. For example, many egalitarians believe that inequality is bad, even when it's not bad *for* anyone. It is easier to see why, on that kind of picture, we should sometimes do what is worse, in expectation, for everyone: the interests of individuals must sometimes give way for the sake of equality. But Buchak explicitly rejects such an egalitarian interpretation of RWU. She motivates RWU with the suggestion that "the only objects of concern are the interests of each individual, but each individual's interests needn't be given the same weight in the evaluation of a distribution" (p. 624). And her argument for RWU appeals to a form of reasoning that is addressed to the potential claims and complaints of each person. Given that motivation, it seems unattractive to violate all individuals' rational preferences.

Rather than rejecting ex ante Pareto, it seems more consistent with Buchak's argument to reject the ex post method of applying RWU under uncertainty. Perhaps, as Rawls argues, we should not "adopt for society as a whole the principle of rational choice for one man" (Rawls 1999, 24). To do so would, arguably, ignore the morally significant difference between intrapersonal and interpersonal tradeoffs. On this line of thought, we cannot infer that society ought to maximize risk-weighted expected moral value from the view that individuals ought to maximize their REU.

If we reject ex post RWU, how should a rank-weighted utilitarian choose between prospects like *C* and *D*?

3   Ex Ante RWU

According to

Ex Ante RWU: We ought to choose one prospect rather than another just in case the one offers a greater rank-weighted average of risk-weighted expected utilities.

The idea behind ex ante RWU is that, under uncertainty, we evaluate distributions of REUs, rather than first evaluating the distributive outcomes and then calculating their risk-weighted expected values. We try to distribute individual risk-weighted expected utilities in a rank-weighted utilitarian manner.

Ex ante RWU satisfies ex ante Pareto. But it violates a weak constraint on social rationality.

Consider table 5. Here we can choose either prospect D from table 4, or prospect E, in which Cat would be worse off than Dan in state 1 $(10, 20)$, but much better off than Dan in state 2 $(50, 20)$.

Table 5: Ex ante RWU violates statewise dominance

| | State 1 (0.5) | | | State 2 (0.5) | | | REU | | RW- |
	Cat	Dan	RWU	Cat	Dan	RWU	Cat	Dan	REU
Prospect D	20	15	16.25	60	20	30	30	16.25	19.6875
Prospect E	10	20	12.5	50	20	27.5	20	20	20

According to ex ante RWU, we ought to choose E. This is because E offers a better distribution of rank-weighted expected utilities, by rank-weighted utilitarian lights. This evaluation is shown in the rightmost column of table 5 under "RW-REU."[9]

It seems to me, however, that we ought to choose D rather than E. For suppose we choose E and learn that state 2 obtains. We would then prefer that we had chosen D, since it would have yielded a better distribution—not just by rank-weighted utilitarian lights, but also according to any plausible theory of distribution. (This is because, in state 2, D *Pareto-dominates* E: in state 2, D makes Cat better off than she would be in E, and makes Dan at least as well off as he would be in E.) Suppose we learn that state 1 obtains. We would then still prefer that we had chosen D, since it would have yielded a better outcome—again, according to any plausible theory of distribution. (This is because, in state 1, D *anonymously* Pareto-dominates E: D makes Cat at least as well off as Dan would be in E, and makes Dan better off than Cat would be in E.) According to

---

[9]D's rank-weighted average of expected utilities is $0.25[0.25(60) + 0.75(20)] + 0.75[0.25(20) + 0.75(15)] = 19.6875$. E's is $0.25[0.25(50) + 0.75(10)] + 0.75[0.25(20) + 0.75(20)] = 20$.

Statewise Dominance: If, in every state of the world with nonzero probability, prospect X yields a better outcome than prospect Y, then we ought to choose X rather than Y.

D statewise dominates E: it is preferable in every state of the world. So, how could it be rational to choose E, when we know that D would guarantee a preferable distribution no matter what? The violation of statewise dominance seems to me an unattractive implication for a view that ultimately appeals to a theory of rational choice (REU theory) that requires statewise dominance.

To emphasize the counterintuitiveness of this result, suppose that we can learn which state the world is in by turning over and looking at a napkin. And suppose we know that we are rank-weighted utilitarians, so that if we look at this napkin, we will certainly choose the best distribution by rank-weighted utilitarian lights.[10] Looking at the napkin would lead us to choose D, so looking offers the same distribution of risk-weighted expected utilities as D. Ex ante RWU therefore implies that we ought not look at the napkin; we should instead just choose E without looking at the napkin. This aversion to costless information, and to doing what that information would lead us to do, seems highly counterintuitive (cf. Fleurbaey and Voorhoeve 2013). I don't think we *have* to look at the napkin, because we know that doing so would lead us to choose D no matter what; but it doesn't seem *wrong* to look at it. The result is even more counterintuitive if we imagine that looking at the napkin would provide some very small benefit to each person. If the benefit to each person is small enough, ex ante RWU will still imply that we ought not look at the napkin, because doing so would lead us to choose a slightly improved version of D that remains worse than E by ex ante RWU's lights. Perhaps there are cases in which we should not acquire costless information (e.g., when that information is private or confidential, or is only partial and might be misleading). But this doesn't seem to be one of them.

We need an argument for ex ante RWU that is more compelling than statewise dominance. An ex ante rank-weighted utilitarian might respond by appealing to the veil of ignorance. The impartial observer is supposed to imagine that they are either Cat or Dan, but don't know which. If they turn out to be Cat, they would prefer D, because their REU is greater under D. If they turn out to be Dan, they would prefer E, because their REU is greater under E. And, according to Buchak, the impartial observer is and ought to be risk-averse, so they give greater weight to the badness of turning out to be Dan under D. The impartial

---

[10] If one has worries about the diachronic aspect of this choice problem, we could instead imagine that we can hand over the choice to another deliberator who is fully informed, or acquire the information in a way that initiates a causal process that results in the best distribution.

observer would therefore rationally prefer *E*.

We should reject this argument. The veil of ignorance does not support the ex ante approach. To explain why, I will raise two preliminary questions about the argument just given; the explanation will emerge from those questions.

First, why should the impartial observer be risk averse when evaluating distributions of *already* risk-weighted expectations? They have already taken the risk for each individual into account when calculating the risk-weighted expected utilities for each person, relative to the most risk-averse attitude within reason. Weighting lower risk-weighted expectations more heavily still seems to double-count individuals' aversion to risk. Perhaps this method can be justified. But how?

In reply, we might point out that the impartial observer is uncertain about two independent questions: which state the world is in, and who they are. The first question is about *wordly* matters; the second is about *self-locating* matters. Since the impartial observer is uncertain about both matters, this gives rise to two kinds of risk: the worldly risk of being worse off than one might have been (e.g., Cat's being worse off in state 1 than in state 2), and the self-locating risk of being a person who is worse off than another (e.g., Dan's being worse off, ex ante, than Cat). If the impartial observer is risk averse, we should count both their aversion to worldly risk and their aversion to self-locating risk.

This reply raises a second question. The veil-of-ignorance argument for ex ante RWU assumes that the impartial observer *first* evaluates the prospects for each person relative to their worldy uncertainty, and *then* evaluates the prospect of being one of those people relative to their self-locating uncertainty. But why should the impartial observer evaluate these two kinds of risk in that particular order? If they instead evaluated the self-locating risk before the worldly risk, we would obtain a very different view. The impartial observer would first assign a value to each outcome in light of their uncertainty about which person they would be in that outcome, and then evaluate the prospect in light of their uncertainty about the state of the world. This procedure would yield ex post RWU.

I see no reason to prefer either of these methods to the other. Without any such reason, we have no reason to think that the impartial observer would choose ex ante RWU, as opposed to ex post RWU. In fact, they would choose neither.

4   Veiled RWU

Rather than evaluating their wordly risk and self-locating risk in two stages, the impartial observer would instead use a simpler, one-stage method: they would treat their two kind of risk on a par.

The impartial observer assigns equal probability to four distinct possibilities: being Cat in state 1, being Dan in state 1, being Cat in state 2, and being Dan in state 2. Call this the *veiled* probability distribution. The impartial observer would maximize their REU relative to the veiled probability distribution because they care only about the rank-ordered distribution of their own outcomes and the probability with which those outcomes are realized. They have no reason to privilege some other way of evaluating their prospects.

Indeed, in the decision-theoretic framework in which Buchak axiomatizes and formulates REU theory, this is the only procedure that makes sense. Prospects are supposed to be functions from states to outcomes. These states have to be fine-grained enough that each prospect assigns exactly one outcome to each state. The impartial observer is uncertain about their identity, so a choice of prospect and wordly proposition does not determine a unique outcome for them; they remain uncertain about which person they are. For the impartial observer, what we have been calling "state 1" and "state 2" are technically not states, strictly conceived, but *events*: sets of states. This distinction is important because, under REU theory, certain kinds of reasoning about states can fail when applied to events—e.g., dominance reasoning: a prospect that is better for you in every state offers you greater REU, but a prospect that is better for you conditional on each event in some arbitrary partition may offer you *lower* REU (Buchak 2013, ch. 5).

To apply REU theory, the impartial observer needs to use the proper partition of states. Table 6 shows the two prospects using the veiled partition above, where "event 1" and "event 2" are just the worldly propositions we had been calling "state 1" and "state 2." Each event has two states, which differ with respect to the impartial observer's identity. Under the veiled partition, prospect D offers the impartial observer greater REU than E.[11]

Using the veiled partition, REU maximization yields neither ex post nor ex ante RWU. It yields

Veiled RWU: We ought to choose one prospect rather than another just in case the one

---

[11]There are four possibilities, so the weights are 1, 3, 5, and 7. The states are equally probable, so we simply apply the normalized weights 1/16, 3/16, 5/16, 7/16. Prospect D has an REU of $1/16(60) + 3/16(20) + 5/16(20) + 7/16(15) = 20.3125$. E has $1/16(50) + 3/16(20) + 5/16(20) + 7/16(10) = 17.5$.

Table 6: Veiled rank-weighted utilitarian evaluation of table 5

	Cat in Event 1 (0.25)	Dan in Event 1 (0.25)	Cat in Event 2 (0.25)	Dan in Event 2 (0.25)	REU
Prospect D	20	15	60	20	20.3125
Prospect E	10	20	50	20	17.5

offers greater REU, relative to the appropriate veiled probability distribution and the most risk-avoidant attitude within reason.[12]

Given a probability distribution over worldly states, the *appropriate* veiled probability distribution preserves the same probability of each corresponding worldly event and assigns a uniform probability distribution over all self-locating events (i.e., being any given person). For example, if we, in the actual world, assign probability p to worldly state 1 and $1 - p$ to worldly state 2, and if there are n people, the appropriate veiled probability distribution assigns p/n to the state of being any given person in worldly event 1, and $(1 - p)/n$ to the state of being any given person in worldly event 2.

Let us consider the verdicts of veiled RWU in the cases we have considered. In table 5, it implies that D is better than E. And, in table 4, it implies that C is better than D. So far, so good.

In other cases, however, veiled RWU can recommend prospects that are worse for everyone *and* prospects that are worse in every state. For example, consider the prospects in table 7. Imagine that we—in the actual world, who know our identities—are choosing between these prospects. We assign an equal probability to each of two worldly states.

Table 7: Veiled RWU violates statewise dominance and ex ante Pareto

	State 1 (0.5)			State 2 (0.5)			REU		Veiled
	Cat	Dan	RWU	Cat	Dan	RWU	Cat	Dan	RWU
Prospect C	20	20	20	50	10	20	27.5	12.5	17.5
Prospect E	10	20	12.5	50	20	27.5	20	20	17.5
Prospect F	46	10	19	10	46	19	19	19	19
Prospect G	46	10	19	46	10	19	46	10	19

---

[12]I take the name from what Thomas (2016 ch. 3) calls "veiled average utilitarianism", which faces a problem similar to the one I raise for veiled RWU (see Nebel 2019, 339, fn. 41).

According to veiled RWU, *F* and *G* are equally good, and would be better than both *C* and *E*. These evaluations are shown under the "Veiled RWU" column.[13] Intuitively, this is because veiled RWU ignores distinctions between both states and persons. It evaluates each person's utility in each outcome as if they were the possible utilities of a single person: the impartial observer. So, given a uniform probability distribution over worldly states, a veiled rank-weighted utilitarian is indifferent between prospects that assign the same list of utilities (e.g., *F* and *G*), regardless of how they are distributed between states and persons.

Veiled RWU implies that, in the actual world—where we know who is who—we ought to choose *F* or *G* rather than *C* or *E*. But note that *C* yields a better distribution than *F* and *G* in every state, and *E* offers greater REU for every person than *F*. Veiled RWU therefore violates both statewise dominance *and* ex ante Pareto. This seems to me highly unattractive.

5 Separability and the Veil of Ignorance

We have considered three ways of applying RWU under uncertainty. The ex post method violates ex ante Pareto. The ex ante method violates statewise dominance. The veiled method violates both principles.

However, we have only seen that these methods *can* violate these principles, because we assumed a particular choice of weighting function. And we have not considered every possible way of applying RWU under uncertainty. In this section, we will see the results of the previous sections in greater generality. This will help us understand, at a deeper level, why the problems arise, and how they apply beyond RWU.

We will demonstrate three results: one for RWU, one for REU theory, and one for veiled RWU. For the first result, we add to ex ante Pareto that, if everyone would rationally be indifferent between two prospects (still relative to a shared probability distribution and risk attitude), then it is permissible to choose either prospect. For the second, we add to statewise dominance that, if two prospects guarantee equally good outcomes in every state, then it is permissible to choose either prospect. The third and most important result builds on the first two but requires neither of the strengthenings just mentioned.

---

[13] *C* and *E* have the same veiled RWU, calculated in note 11. *F* and *G* also have the same value: $1/16(46) + 3/16(46) + 5/16(10) + 7/16(10) = 19$.

5.1 RWU and Separability of People

Utilitarians (and prioritarians in the style of Parfit (1997)) believe that value is *strongly separable* with respect to people. This means, roughly, that whether or not one distribution is better than another cannot depend on the welfare of groups of people who are equally well off in both distributions. Rank-weighted utilitarians reject this claim, because the importance of benefiting a person can, on their view, depend on how well off other people are. Consider, for example, the four distributions in table 8. In each distribution (row), there are two groups of people (columns). Group 1 has some number m of people; group 2 has n people. Each cell depicts a *subdistribution* over the corresponding group of people. Subdistributions A and B assign various welfare levels to each person in group 1; subdistributions X and Y assign various welfare levels to each person in group 2.

Table 8: Rank-weighted utilitarians reject separability of people

	Group 1 (m)	Group 2 (n)
Distribution AX	A	X
Distribution BX	B	X
Distribution AY	A	Y
Distribution BY	B	Y

Strong separability implies that distribution AX is better than BX if and only if AY is better than BY, for any values of m and n, and any subdistributions A, B, X and Y: group 2 is unaffected by the choice between AX and BX, and by the choice between AY and BY, so these comparisons should depend only on the subdistributions A and B. Rank-weighted utilitarians, by contrast, maintain that for some m and n, and some A, B, X, and Y, AX can be better than BX while BY is better than AY. A concrete example, using the Gini weighting, is depicted in table 9.

Table 9: Violation of strong separability for Gini RWU

	Group 1		Group 2		
	i	j	k	l	RWU
Distribution AX	32	32	32	32	32
Distribution BX	16	64	32	32	27
Distribution AY	32	32	80	80	44
Distribution BY	16	64	80	80	47

In table 9, RWU with the Gini weighting implies that *AX* is better than *BX* but that *BY* is better than *AY*. This is because the presence of the unaffected group affects the distribution's rank ordering. In particular, the greater welfare of person *j* has less weight when *j* is better off than everyone else than when *j* is worse off than others.[14] Though this particular example uses the Gini weighting, every version of RWU implies that, in some case of the kind depicted in table 8, *AX* is better than *BX* but *BY* is better than *AY*.

So suppose that, for some group sizes and subdistributions characterized by table 8, *AX* is better than *BX* but *BY* is better than *AY*. Let "∘" represent the probabilistic mixture of two distributions into a social prospect: *AX* ∘ *BY*, for example, represents a prospect that assigns *AX* to one of two equiprobable states and *BY* to the other. And consider the two prospects in table 10.

Table 10: RWU violates statewise dominance or ex ante Pareto

	State 1 (0.5)		State 2 (0.5)	
	Group 1 (*m*)	Group 2 (*n*)	Group 1 (*m*)	Group 2 (*n*)
Prospect *AX* ∘ *BY*	*A*	*X*	*B*	*Y*
Prospect *BX* ∘ *AY*	*B*	*X*	*A*	*Y*

We are supposing that distribution *AX* is better than *BX* but *BY* is better than *AY*. So, prospect *AX* ∘ *BY* guarantees a better distribution than prospect *BX* ∘ *AY*. So, by statewise dominance, we ought to choose *AX* ∘ *BY* rather than *BX* ∘ *AY*. But everyone in both groups would rationally be indifferent between the two prospects, whatever risk attitude they have. Everyone in group 1 will either have their welfare under *A* or their welfare under *B*, with equal probability of each. Everyone in group 2 will either have their welfare under *X* or their welfare under *Y*, with equal probability of each. These two prospects have the same REU for each person. So, by ex ante Pareto (as strengthened above), it must be permissible to choose either prospect. This shows that RWU, statewise dominance, and ex ante Pareto are jointly inconsistent.

Two things about this result are worth noting. First, it applies to any distributive theory that rejects the separability of people—e.g., standard versions of egalitarianism. Second, the conflict does not assume that individuals are REU maximizers. It only assumes that an

---

[14] *AX* has a rank-weighted average of $1/16(32) + 3/16(32) + 5/16(32) + 7/16(32) = 32$. *BX* has $1/16(64) + 3/16(32) + 5/16(32) + 7/16(16) = 27$. *AY* has $1/16(80) + 3/16(80) + 5/16(32) + 7/16(32) = 44$. *BY* has $1/16(80) + 3/16(80) + 5/16(64) + 7/16(16) = 47$. Person *j*'s weight is increased from 1/16 in *BX* to 5/16 in *BY*.

individual would rationally be indifferent between prospects that assign the same outcomes for her to states of the same probability. We can demonstrate a similar result for REU theory without appealing to RWU. This will help us to establish our third result, for veiled RWU.

5.2 REU and the Sure-Thing Principle

Expected utility theory satisfies the *sure-thing principle*, which is a separability principle for states. According to EU theory, whether or not one prospect is better for a person than another cannot depend on the value of outcomes that both prospects assign to the same set of states. REU theorists rejects this claim, because the importance of an outcome can, on their view, depend on how good other outcomes are. Consider, for example, the prospects in table 11. Here *A*, *B*, *X*, and *Y* are *subprospects*: *A* and *B* assign various outcomes to the states in event 1; *X* and *Y* assign various outcomes to the states in event 2.

Table 11: REU and the sure-thing principle

	Event 1 (p)	Event 2 $(1-p)$
Prospect *AX*	*A*	*X*
Prospect *BX*	*B*	*X*
Prospect *AY*	*A*	*Y*
Prospect *BY*	*B*	*Y*

According to REU theory, it can be rational for an individual to prefer prospect *AX* to *BX* while preferring *BY* to *AY*. Now consider two people who have those same rational preferences, and consider the prospects in table 12. In table 12, both people rationally prefer prospect $AB \circ XY$ to prospect $BA \circ XY$. This is because person 1 prefers *AX* to *BX*, and person 2 prefers *BY* to *AY*. So, by ex ante Pareto, we ought to choose $AB \circ XY$ rather than $BA \circ XY$. This violates statewise dominance (as strengthened above) because the two prospects guarantee equally good distributions: they just switch which person gets which outcome.

Table 12: REU violates statewise dominance or ex ante Pareto

	Event 1 (p)		Event 2 $(1-p)$	
	Person 1	Person 2	Person 1	Person 2
Prospect $AB \circ XY$	*A*	*B*	*X*	*Y*
Prospect $BA \circ XY$	*B*	*A*	*X*	*Y*

This shows that REU theory, statewise dominance, and ex ante Pareto are incompatible—not just given RWU, but more generally if we assume that distributions in which the same number of people are at the same welfare levels are equally good. Note also that this result applies not only to REU theory, but to any decision theory that violates the sure-thing principle—e.g., the prospect theory of Kahneman and Tversky (1979).

5.3   The Impartial Observer

Now suppose that we appeal to the preferences of the impartial observer. This leads to the veiled approach. On the veiled approach, RWU and REU theory are equivalent. So, in table 8, distribution AX is better than BX for group sizes m and n just in case, in table 11, the impartial observer would rationally prefer prospect AX to prospect BX for $p = m/(m + n)$. The same goes for the comparison between AY and BY.

We can now show that veiled RWU violates both statewise dominance and ex ante Pareto (in their weaker original versions).

Start with statewise dominance. To see the violation, consider again prospects $AX \circ BY$ and $BX \circ AY$ in table 10. On the veiled approach, these two prospects should be evaluated relative to a veiled probability distribution, which contains four events. Let "event i, j" denote the event in which worldly state i obtains and the impartial observer belongs to group j. Table 13 compares the prospects relative to the corresponding veiled probability distribution. Veiled $AXBY$ is how the impartial observer evaluates prospect $AX \circ BY$; veiled $BXAY$ is how they evaluate $BX \circ AY$.

Table 13: Veiled RWU violates statewise dominance

	Event 1,1 $0.5(m/[m + n])$	Event 1,2 $0.5(n/[m + n])$	Event 2,1 $0.5(m/[m + n])$	Event 2,2 $0.5(n/[m + n])$
Veiled $AXBY$	A	X	B	Y
Veiled $BXAY$	B	X	A	Y

Veiled $AXBY$ and veiled $BXAY$ offer the exact same REU, so the impartial observer would be indifferent between these two prospects. So, according to veiled RWU, it is permissible to choose either $AX \circ BY$ or $BX \circ AY$ in table 10. But, as we already saw, that violates statewise dominance, which implies that we ought to choose $AX \circ BY$. This shows that veiled RWU violates statewise dominance.

Next consider ex ante Pareto. Consider again prospects $AB \circ XY$ and $BA \circ XY$ in table 12. We should compare these two prospects relative to the appropriate veiled probability distribution. The resulting "veiled" prospects are shown in table 14. Here "event i, j" denotes the event in which worldly event i obtains and the impartial observer is person j.

Table 14: Veiled RWU violates ex ante Pareto

	Event 1,1 $0.5(m/[m+n])$	Event 1,2 $0.5(m/[m+n])$	Event 2,1 $0.5(n/[m+n])$	Event 2,2 $0.5(n/[m+n])$
Veiled $ABXY$	A	B	X	Y
Veiled $BAXY$	B	A	X	Y

According to veiled RWU, we ought to choose prospect $AB \circ XY$ rather than $BA \circ XY$ in table 12 just in case, in table 14, the impartial observer would rationally prefer veiled $ABXY$ to veiled $BAXY$. But they would rationally be indifferent between these two veiled prospects, because they offer the same probabilities of the same outcomes; they must have equal REU, relative to any risk attitude. This shows that veiled RWU violates ex ante Pareto.

Thus, veiled RWU violates both statewise dominance and ex ante Pareto. This result applies, more generally, to any theory that appeals to the veil of ignorance while rejecting either the sure-thing principle or the separability of people.

6   Conclusion

Buchak's argument for RWU appeals to the preferences of rational agents behind the veil of ignorance, who are supposed to maximize their REU. This leads to violations of both statewise dominance and ex ante Pareto. Since we have seen that statewise dominance and ex ante Pareto together impose quite severe constraints on both social and rational choice, perhaps it is reasonable to reject one of these conditions. But, although many of us may be willing to reject one of these conditions, it seems hard to reject them both. These violations are especially troubling for Buchak because REU theory entails statewise dominance, and because Buchak motivates the veil of ignorance by the requirement that our acts be justifiable to each person, which seems to support ex ante Pareto. I do not think this refutes REU theory or RWU. But I think it casts doubt on Buchak's *argument* for RWU, according to which we ought to choose whatever would be preferred by risk-avoidant REU maximizers who do not know who they are.

I myself am inclined to reject both REU theory and RWU for reasons independent of these issues. But the results of this paper provide some reason for fans of these theories—or, more generally, of any nonseparable theories of distribution or decision—not to appeal to the veil of ignorance. The veil of ignorance may be a valuable heuristic device for ensuring impartiality. But, as Parfit puts it, "it does that crudely, like frontal lobotomy" (2011, 357). It requires us to ignore information that may be relevant to distributive justice—i.e., which utilities belong to whom, and in which outcomes. We should not make distributive choices by depriving ourselves of this information, but by ensuring that we are impartial in other ways, if we can (cf. Roemer 2002).

References

Adler, Matthew. 2011. *Well-Being and Fair Distribution: Beyond Cost-Benefit Analysis*. Oxford University Press. http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780195384994.001.0001/acprof-9780195384994.

Blessenohl, Simon. 2020. "Risk Attitudes and Social Choice." *Ethics*.

Buchak, Lara. 2013. *Risk and Rationality*. OUP Oxford. http://books.google.com?id=kBgbAgAAQBAJ.

———. 2017. "Taking Risks Behind the Veil of Ignorance." *Ethics* 127 (3): 610–44. https://doi.org/10.1086/690070.

Diamond, Peter A. 1967. "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparison of Utility: Comment." *Journal of Political Economy* 75 (5): 765–66. https://doi.org/10.1086/259353.

Fleurbaey, Marc, and Alex Voorhoeve. 2013. "Decide as You Would with Full Information!" In *Inequalities in Health*, edited by Nir Eyal, Samia A. Hurst, Ole F. Norheim, and Dan Wikler, 113–28. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199931392.003.0009.

Harsanyi, John C. 1953. "Cardinal Utility in Welfare Economics and in the Theory of Risk-Taking." *Journal of Political Economy* 61. https://ideas.repec.org/a/ucp/jpolec/v61y1953p434.html.

———. 1955. "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility." *Journal of Political Economy* 63 (4): 309–21. https://www.jstor.org/stable/1827128.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision Under Risk." *Econometrica* 47 (2): 263–91. https://doi.org/10.2307/1914185.

Kowalczyk, Kacper. 2019. "Allais Preferences Are Bad for You." Unpublished manuscript. St. Anne's College, Oxford: Unpublished manuscript.

McCarthy, David, Kalle Mikkola, and Teruji Thomas. 2020. "Utilitarianism with and Without Expected Utility." *Journal of Mathematical Economics* 87 (March): 77–113. https://doi.org/10.1016/j.jmateco.2020.01.001.

Mongin, Philippe, and Marcus Pivato. 2015. "Ranking Multidimensional Alternatives and Uncertain Prospects." *Journal of Economic Theory* 157 (May): 146–71. https://doi.org/10.1016/j.jet.2014.12.013.

Nebel, Jacob M. 2019. "An Intrapersonal Addition Paradox." *Ethics* 129 (2): 309–43. https://doi.org/10.1086/700085.

Parfit, Derek. 1997. "Equality and Priority." *Ratio* 10 (3): 202–21.

———. 2011. *On What Matters. V. 1*. Oxford: Oxford University Press.

Rawls, John. 1999. *A Theory of Justice*. Cambridge, Mass.: Belknap Press of Harvard University Press.

Roemer, John E. 2002. "Egalitarianism Against the Veil of Ignorance." *The Journal of Philosophy* 99 (4): 167–84. https://doi.org/10.2307/3655614.

Thomas, Teruji. 2016. "Topics in Population Ethics." D.Phil. Thesis, University of Oxford.

Weymark, John A. 1981. "Generalized Gini Inequality Indices." *Mathematical Social Sciences* 1 (4): 409–30. https://doi.org/10.1016/0165-4896(81)90018-4.