

# Respect for others' risk attitudes and the long-run future

Andreas L. Mogensen

University of Oxford, Oxford, UK

## Correspondence

Andreas L. Mogensen, University of Oxford, Oxford, UK.

Email:

[andreas.mogensen@philosophy.ox.ac.uk](mailto:andreas.mogensen@philosophy.ox.ac.uk)

## Abstract

When our choice affects some other person and the outcome is unknown, it has been argued that we should defer to their risk attitude, if known, or else default to use of a risk-avoidant risk function. This, in turn, has been claimed to require the use of a risk-avoidant risk function when making decisions that primarily affect future people, and to decrease the desirability of efforts to prevent human extinction, owing to the significant risks associated with continued human survival. I raise objections to the claim that respect for others' risk attitudes requires risk-avoidance when choosing for future generations. In particular, I argue that there is no known principle of interpersonal aggregation that yields acceptable results in variable population contexts and is consistent with a plausible ideal of respect for others' risk attitudes in fixed population cases.

## 1 | INTRODUCTION

According to orthodox decision theory, a rational agent in conditions of uncertainty prefers those acts that maximize expected utility. In the recent philosophical literature, an influential alternative to expected utility theory is defended by Buchak (2013), who argues for the rationality of maximizing *risk-weighted expected utility* (REU). On this view, a rational agent's preferences over uncertain prospects depend not only on the probabilities she assigns to the different possible

---

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Authors. *Noûs* published by Wiley Periodicals LLC.

states of the world and the utility of the different possible outcomes, but also independently on her attitude toward risk, as captured by a risk function on probabilities.

More recently, Buchak (2016, 2017, 2019) has argued that other-regarding contexts require that we exhibit a high degree of risk-avoidance as a default. This default, she claims, should especially guide those of our decisions whose largest impacts are on future individuals, such as decisions about climate change. In a recent paper, Pettigrew (2022b) argues that the requirement to default to risk-avoidance for decisions whose largest impacts are on future individuals pushes total utilitarians in the direction of favouring premature human extinction over continued human survival, in light of the significant risks associated with the persistence of human beings as a dominant species.

The core idea that motivates these lines of argument is that of *respect for others' risk attitudes*. When our actions affect others, we ought not simply impose our own idiosyncratic attitude toward risk on them, and should instead choose in a way that takes account of – and can be justified relative to – the possible risk attitudes of the people we potentially affect. In particular, we ought to make it the case that our choice cannot reasonably be judged as too risky by the people we affect, given what we know about their risk attitudes.

I will offer reasons to think that a plausible ideal of respect for others' risk attitudes does not support the kind of conclusions outlined by Buchak and Pettigrew, and may even be irrelevant when thinking about the impact of present actions on the long-run future. The kind of principle that seems to be justified by an ideal of respect for others' risk attitudes in fixed population cases is one on which we ought to promote the aggregate of each individual's risk-weighted expected welfare, and not the risk-weighted expectation of aggregate welfare. For gambles where some people exist in only some outcomes and not in others, it makes no sense to speak of those individuals' risk-weighted expected welfare under the gamble, nor the aggregate of each person's risk-weighted expected welfare. Consequently, we cannot be guided by the kind of principle supported by a plausible ideal of respect for others' risk attitudes when making decisions among uncertain options in variable population contexts.

I begin in section 2 by outlining the theory of risk-weighted expected utility. In section 3, I then set out Buchak's view that we should default to a risk-avoidant risk attitude when making choices on behalf of others, and I outline claims made by Buchak and Pettigrew about the implications of this principle for actions whose most important effects concern future people. In section 4, I emphasize the choice we face when trying to respect others' risk attitudes when choosing for a group of persons, between aggregating first across persons and then across outcomes, or *vice versa*. I note that there is good reason to think that respect for others' risk attitudes requires that we adopt the latter approach, but that this approach is incompatible with consequentialism. In section 5, I argue, along the lines noted above, that this approach breaks down in variable population cases of the kind we inevitably confront in making decisions about the long-run future, and I discuss possible responses. Section 6 provides a summary and conclusion.

## 2 | RISK-WEIGHTED EXPECTED UTILITY THEORY

I first explain the idea of risk-weighted expected utility and the corresponding notion of risk-avoidance.

Assume there are  $n$  possible states of the world,  $S_1, \dots, S_n$ . If act  $f$  is chosen, corresponding outcomes  $x_1, \dots, x_n$  will be realized depending on which of  $S_1, \dots, S_n$  is actual. We assume the agent's beliefs about which state is actual are represented by a subjective probability function,

$\Pr(\cdot)$ , and the agent's preferences over outcomes are represented by a cardinal utility function,  $u(\cdot)$ . We also assume that states and outcomes are indexed in ascending preference order, such that  $u(x_i) \leq u(x_{i+1})$  for  $i = 1, \dots, n - 1$ . Then the risk-weighted expected utility of  $f$  is defined as

$$REU(f) := u(x_1) + \sum_{k=2}^n \left[ r \left( \sum_{i=k}^n \Pr(S_i) \right) \cdot (u(x_k) - u(x_{k-1})) \right]$$

In other words, we add to the utility of the worst possible outcome,  $x_1$ , the gain in utility provided by instead obtaining the next worst outcome,  $x_2$ , weighted by applying a function,  $r(\cdot)$ , to the probability of attaining an outcome at least as good as  $x_2$ ; and then we add the gain in utility provided by moving up from the second-to-last-ranked outcome to the third-to-last-ranked outcome,  $x_3$ , weighted by  $r(\cdot)$  of the probability of attaining an outcome at least as good as  $x_3$ , and so on.

We call  $r(\cdot)$  the *risk function* and stipulate that it is a nondecreasing function that maps the unit interval to itself and has 0 and 1 as fixed points. As its name suggests, the risk function is intended to encode the agent's attitude toward risk. If  $r(\cdot)$  is linear, then risk-weighted expected utility reduces to expected utility. According to Buchak, we are not rationally required to have a linear risk function. When  $r(\cdot)$  is strictly convex — e.g.,  $r(\Pr(X)) = \Pr(X)^2$  — the agent is said to be *risk-avoidant* and puts more weight on the worst possible outcomes of a gamble.

REU theory can explain some cases of apparently rational decision-making under conditions of risk that cannot be captured by expected utility theory. Consider, in particular, the famous *Allais paradox* (Allais, 1953), to which we'll return again later. Imagine a lottery in which there are 100 numbered tickets and four possible gambles grouped into two pairs, yielding different payoffs depending on the number of the ticket drawn at random:

TABLE 1 The Allais Paradox.

	Ticket #1-89	Ticket #90-99	Ticket #100
Gamble A	\$1M	\$5M	\$0
Gamble B	\$1M	\$1M	\$1M
Gamble C	\$0	\$5M	\$0
Gamble D	\$0	\$1M	\$1M

In choices of this kind, people tend to prefer B to A but also prefer C to D (Slovic & Tversky, 1974). These preferences seem inconsistent in principle with valuing an uncertain prospect at its expected utility,<sup>1</sup> but can be accommodated by the assumption that agents maximize risk-weighted expected utility and are risk-avoidant.<sup>2</sup>

<sup>1</sup> A suggestion going back to Raiffa (1968) is that the Allais preferences can be made consistent with expected utility theory by redescribing the possible outcomes to take account of additional concerns that agents may have beyond the monetary prizes themselves, such as the avoidance of regret. See Baccelli and Mongin (2021) for discussion. More recently, Stefánsson and Bradley (2019) have argued that the Allais preferences are consistent with expected utility maximization if the domain of the utility function is enlarged to include chance propositions.

<sup>2</sup> But see Thoma and Weisberg (2017).

### 3 | RESPECT FOR OTHERS' RISK ATTITUDES

It may be thought that if a wide range of risk functions are rationally permissible, then different agents can differ significantly in their evaluation of acts whose most important potential effects concern future individuals, even if they agree in their moral outlook and their empirical beliefs, by virtue of assessing those acts according to their own idiosyncratic risk function. Buchak (2019) argues for a notably different view. According to her *Future Risk-Avoidance Principle* (FRAP), “If we are making a decision whose largest effects concern a large group of future individuals, then we should make a very risk-avoidant choice: a choice which weights the worse consequences proportionally much more heavily than the better consequences.” (Buchak, 2019, 76) Morally speaking, the agent’s own idiosyncratic risk function must be set aside – unless it happens to be very risk-avoidant.

How does she arrive at this conclusion? According to Buchak (2016, 2017, 2019) we are morally required to choose in accordance with a risk attitude that is sensitive to the risk attitudes of the agents potentially affected by our decision. We ought not simply impose our own risk attitude on others. Moreover, when the risk attitudes of those we affect through our actions are unknown to us, we are morally required, she claims, to default to the most risk-avoidant risk attitude within reason,<sup>3</sup> so that our choice cannot reasonably be rejected as excessively risky by those we might affect.<sup>4</sup>

These commitments are encapsulated in Buchak’s *Risk Principle* (RP): “When making a decision for an individual, choose under the assumption that he has the most risk-avoidant attitude within reason unless we know that he has a different risk attitude, in which case, choose using his risk attitude.” (Buchak, 2017, 632)<sup>5</sup> Buchak then derives FRAP from RP, as follows. Since the risk attitudes of future people are unknown to us, insofar as our actions affect future individuals, those actions should be governed by the most risk-avoidant risk attitude within reason. Therefore, for any reasonable principle for aggregating across groups of individuals with diverse risk attitudes,

<sup>3</sup> Note that Buchak here distinguishes between those risk attitudes that are *rational*, in the sense of satisfying certain minimal coherence constraints, and those attitudes that are *reasonable*, in the sense of satisfying certain additional substantive normative constraints (compare Rawls, 1993). In this sense, it may not be irrational to prefer the destruction of the whole world to the scratching of your finger, but it is unreasonable. The most risk-avoidant risk attitude within reason is taken to be the extremal member of the set of reasonable risk attitudes, and not the set of rational risk attitudes.

<sup>4</sup> Clearly, it is difficult to say exactly what kind of risk-avoidant attitude is within the outer bounds of reasonableness, but to give the reader some sense of this, Buchak (2019, 73) suggests that “it is not unreasonable to care about the bottom half of consequences five times as much as the top half, but that is close to the reasonable lower limit.” Thus, consider a gamble over outcomes  $x$ , and  $y$ , where each has a .5 probability and  $u(x) > u(y)$ . Then, according to Buchak, it is not unreasonable to have a risk function,  $r$ , such that  $r(1) - r(.5) = 5 \cdot r(.5)$ , or, equivalently, such that  $r(.5) = 1/6$ . Nonetheless, this is close to the outer bounds of reasonableness. Thanks to Lara Buchak for help in clarifying this example.

<sup>5</sup> How should we understand talk of “his risk attitude”? This way of speaking may seem to presuppose that there exists a single risk function that the agent uses for decision-making in all domains. Buchak (2013) tentatively adopts this assumption. However, principles she later proposes, like RP and FRAP, seem to presuppose that agents may use a different risk function when deciding for other people than when deciding on their own behalf. When Buchak speaks of the agent’s own attitude toward risk, I shall take her to mean the risk function they would use in assessing uncertain prospects apart from any normative constraints extrinsic to the requirements of instrumental rationality, like RP and FRAP. I assume for simplicity that there is a unique risk function satisfying this description and that the agent’s risk attitude does not otherwise exhibit domain dependence. For agents that violate this assumption, it would be straightforward to modify the discussion by framing matters in terms of the individual’s attitude toward risk in relation to the outcome domain relevant to the given decision problem. Thanks to an anonymous referee for pressing me to clarify this point and to Lara Buchak for help in addressing this issue.

we should expect that decisions that primarily affect future people are required to be governed by a highly risk-avoidant risk attitude. Buchak argues on this basis that climate policies ought to conform closely to the kind of recommendations typically thought to follow from the Precautionary Principle (Steel, 2015).

More recently, Pettigrew (2022b) has argued that (a suitably refined version of) FRAP undermines the case for thinking that total utilitarianism requires us to minimize the risk of near-term human extinction (Bostrom, 2003) and instead suggests we ought to work toward our species' end. The survival of humanity, after all, is a kind of gamble. It could go well, and it could go terribly wrong. Pettigrew shows that, using apparently reasonable probability and value assignments for different possible long-run outcomes in a simplified model, if we are morally required to choose in accordance with a highly risk-avoidant risk attitude, then we should prefer to support initiatives to bring about voluntary human extinction insofar as our goal is to maximize total welfare. Pettigrew takes the availability of this kind of argument to cast doubt on the plausibility of setting priorities by combining theoretical results from population axiology and decision theory and following the argument where it leads, the form of argument he takes to be associated with those making the case for *strong longtermism* – roughly, the view that “far-future effects are the most important determinant of the value of our options” (Greaves & MacAskill, 2021, 3).

#### 4 | INTERPERSONAL AGGREGATION AND CONSEQUENTIALISM

Of itself, RP governs only the case of making a decision for an individual. However, decisions about climate change and existential risk potentially impact large groups of individuals. How should we decide in a manner that is respectful of others' risk attitudes when our actions potentially affect many different people in different ways? Buchak asserts that any reasonable principle of interpersonal aggregation in this context vindicates FRAP. Is she right?

In making the case for answering in the negative, I want to focus our attention on the following issue. When deciding for a group of persons in a way that respects others' risk attitudes, one option is to aggregate people's risk functions so as to arrive at a collective risk function, and to choose the act that maximizes the risk-weighted expectation of aggregate welfare relative to the collective risk function. When everyone has the same risk function, we can assume it serves as the collective risk function, so understood. A different possibility is instead to rank the actions available to us by aggregating people's risk-weighted expected welfare relative to their own personal risk function or the risk function we impute to them.

Suppose, for example, that we are attracted to an additive principle for interpersonal aggregation. When people all have or are imputed to have the same risk function,  $r(\cdot)$ , we then face the choice of whether to prefer the act that maximizes the risk-weighted expectation of the sum of each person's welfare relative to  $r(\cdot)$  or the sum of each person's risk-weighted expected welfare relative to  $r(\cdot)$ . Given a fixed population and a fixed probability function, there is no difference between the sum of each person's expected welfare and the expectation of the sum of each person's welfare, so these coincide when  $r(\cdot)$  is linear. The same is not true when a non-linear risk function is applied (Blessenohl, 2020).

To see this, consider the following case, (compare Nebel, 2021, 103–104, Bradley, 2022, 561–563). Imagine that both Afryea and Beom-seok have the same level of wealth and the same function converting dollars to welfare. They also have the same convex risk function (e.g.,  $r(\Pr(X)) = \Pr(X)^2$ ), and in maximizing their risk-weighted expected welfare, each exhibits the Allais preferences: each prefers B to A and C to D in Table 1. Imagine that we have the following choice:

TABLE 2 The Allais Paradox rides again.

	Ticket #1-89	Ticket #90-99	Ticket #100
Gamble E	Afryea: \$1M	Afryea: \$5M	Afryea: \$0
	Beom-seok: \$0	Beom-seok: \$1M	Beom-seok: \$1M
Gamble F	Afryea: \$1M	Afryea: \$1M	Afryea: \$1M
	Beom-seok: \$0	Beom-seok: \$5M	Beom-seok: \$0

If you compare these gambles to those in the Allais Paradox (Table 1), you'll see that Gamble F here corresponds to assigning gamble B to Afryea and gamble C to Beom-seok, whereas Gamble E corresponds to assigning gamble A to Afryea and gamble D to Beom-seok. Since both Afryea and Beom-seok have greater risk-weighted expected welfare under Gamble B than A and under Gamble C than D in Table 1, in Table 2, F uniquely maximizes the sum of each individual's risk-weighted expected welfare. However, since they have the same welfare function on dollars and the outcomes of the possible gambles simply permute the winnings of the different individuals within states, the sum of their welfare is guaranteed to be the same regardless of what we choose, and so E and F have the same risk-weighted expectation of total welfare.

The contrast we have just observed clearly does not require us to presuppose a summative principle of aggregation. F uniquely maximizes the aggregate of each individual's risk-weighted expected welfare given any principle for aggregating risk-weighted expected welfare satisfying *Ex Ante Pareto*, which says that if one gamble has greater risk-weighted expected welfare for each person, then it has greater aggregate risk-weighted expected welfare overall. Aggregate welfare is guaranteed to be the same regardless of whether we choose E or F assuming any principle for aggregating welfare within outcomes that satisfies *Anonymity*, which says that if two outcomes have the same number of people at the same welfare levels then they have the same aggregate welfare, all else being equal. Given *Ex Ante Pareto* and *Anonymity*, maximizing the aggregate of each individual's risk-weighted expected welfare renders choice of E impermissible, whereas maximizing the risk-weighted expectation of the aggregate of each person's welfare permits choice of E (Nebel, 2021; compare Blessenohl, 2020 and Bradley, 2022).

Besides the fact that the different approaches to aggregation yield different results, Table 2 also illustrates the following point. Given any theory of the good on which outcomes with the same aggregate welfare are equally good and the welfare aggregation function satisfies *Anonymity*, insofar as our goal is to produce morally good outcomes, we presumably ought to be indifferent between E and F. To derive this result we need only assume a *weak dominance principle*, on which, insofar as our goal is to produce morally good outcomes, one gamble is not strictly preferable to another if the latter yields an outcome that is morally at least as good in every possible state of the world. Since the weak dominance principle is extremely plausible, and so is the assumption that the right theory of the good counts two outcomes as equally good that have the same number of people at the same welfare levels, all else being equal, it is highly plausible that any view on which choice of F is strictly preferable to E must be concerned with something other than the value of outcomes. Therefore, choosing on the basis of the aggregate of each individual's risk-weighted expected welfare is inconsistent with consequentialism, understood as the class of moral views according to which the moral value of outcomes is the only determinant of what we morally ought to choose.

What does RP prescribe in cases like that illustrated in Table 2? Since it is formulated in a way that only applies to the case of choosing for an individual, strictly speaking, it is silent. Nonetheless, the considerations that motivate RP seem to me to strongly favour strictly preferring F to E. At

least in Buchak's presentation, RP is motivated by the thought that when choosing for others, we should err against subjecting people to risks we're not sure they would take on their own behalf. Thus, Buchak holds that "we cannot choose a more-than-minimally-risky gamble for another person unless we have some reason to think that he would take that gamble himself" (Buchak, 2019, 74). The ideal of justifiability to each individual is also taken to support RC, in the form of the idea that we should "take only the risks that no one could reasonably reject." (ibid.) In the example above, both Afryea and Beom-seok would prefer F to E if choosing rationally on their own behalf, and choice of F is, in that sense, uniquely justifiable to each in light of her risk attitude (compare Frick, 2015, 186–191). Therefore, the considerations that motivate RP seem to favour choice of the option that tracks the aggregate of each individual's risk-weighted expected welfare, rather than the risk-weighted expectation of the aggregate of each individual's welfare.

For the reasons already noted, this arguably cannot be made consistent with choosing among options based on exclusive concern for achieving morally good outcomes. But nor should this be especially surprising. As just noted, RP is supported by the ideal of justifiability to each person. This ideal is most strongly associated with the Kantian contractualist moral theory developed by Scanlon (1998). More generally, it is unclear what recommends deferring to people's risk attitudes if all that matters morally is achieving morally good outcomes and people's risk attitudes are not concerned with the value of outcomes for them, as on Buchak's theory (Thoma, 2023). It is much easier to motivate on a broadly Kantian conception of the moral domain on which respect for the autonomous personhood of others plays a crucial part in determining our conception of right and wrong, and it is natural to understand RP as related to ideals of respect for another person's sovereignty in matters that concern her well-being similar to those that figure in debates on paternalism.

The definition of paternalism is contested, but on at least one plausible account, I act paternalistically toward another agent if I act for her sake but do not treat her will as authoritative in determining what I should do for her (Groll, 2012).<sup>6</sup> If I were to choose on behalf of another but in accordance with my own attitude toward risk rather than hers, I would be acting in just this way. I would be substituting my own attitude for hers in determining what is best for her. Many authors require that paternalism involves a judgment that the paternalized subject is impaired in her capacity to promote her own welfare (Enoch, 2016; Quong, 2010; Shiffrin, 2000). This need not occur in the kind of examples we are considering, since the different risk attitudes of the actor and the beneficiary may be acknowledged as equally rational and reasonable. While this may count against applying the label 'paternalism,' a core aspect of the wrong in paternalism arguably remains: namely, a failure to respect another's rational will as authoritative when acting for her sake, by substituting one's own judgment for hers. While the absence of potentially patronizing attitudes on behalf of the agent may make her conduct easier to swallow in one sense, viewed in another, the decision to substitute the paternalizing agent's judgment for her intended beneficiary's is surely even harder to justify if the paternalizing agent makes no assumption that she knows better.

The points I've gone over are especially important for assessing Pettigrew's argument. Pettigrew's approach to choosing on behalf of groups of individuals is the contrary of the approach

<sup>6</sup> Note, however, that Groll (2012) ultimately hesitates in adopting this particular account of paternalism, in preference to a definition on which the paternalizing agent must judge the paternalized subject as likely to fail in exercising a capacity for sound judgment. Nonetheless, he notes he is unsure whether the latter ought to be viewed as complementing or supplanting a definition of the kind noted above, and he acknowledges that it "offers a less intuitive account of the impermissibility of paternalism when the paternalized subject clearly has a right to make the choice in question." (Groll, 2012, 717 n.51)

I have claimed is recommended to us insofar as we find RP plausible. Pettigrew (2022b, 23–24) recommends that “[w]hen you make a decision on behalf of a group of people that might result in harm to the people in that group, you should use a risk attitude obtained by aggregating the risk attitudes that those people have. And, when performing this aggregation, you should give greater weight to more risk-averse individuals in the group.” Thus, if there are  $n$  individuals with risk functions  $r_1(\cdot), \dots, r_n(\cdot)$ , Pettigrew suggests that the collective risk function  $r_G(\cdot)$  might be obtained by taking a weighted average,  $r_G(\cdot) = \lambda_1 r_1(\cdot) + \dots + \lambda_n r_n(\cdot)$ , where the weights sum to 1 and  $\lambda_i$  is greater the more risk-avoidant  $r_i(\cdot)$  is. This collective risk function is then to be used in assessing different options in terms of the risk-weighted expectation of the values assigned to the different possible outcomes by aggregating welfare within those outcomes. For the reasons already noted, this approach yields conclusions that seem difficult to square with a plausible ideal of respect for others’ risk attitudes.

Pettigrew (2022b, 32–4) does address the objection that his preferred decision procedure permits us to go against the unanimous self-regarding preferences of the people affected by our decision in cases similar to that presented in Table 2. He replies that this should not worry us, since any plausible decision theory may permit – and also require – us to choose an option that violates the unanimous self-regarding preferences of the affected parties if we aim to maximize the good. Even when individuals’ self-regarding preferences over gambles are based on their expected welfare, the goal of maximizing total welfare can permit and require us to go against the unanimous self-regarding preferences of the affected parties in conditions of uncertainty if the parties differ in their subjective probabilities (Broome, 1987), as may be thought rationally permissible even if they have the same evidence (Kelly, 2013; Pettigrew, 2022a; Schoenfield, 2014) (compare Blessenohl, 2020, 499–501).

However, the objection I am raising is not that we ought not go against the unanimous preference of the affected parties in a case like that presented in Table 2. To raise that objection would be to presuppose the falsity of consequentialism as a theory of right action given any welfarist axiology satisfying Anonymity. My objection is intended as neutral with respect to the correct moral theory. My complaint is that a moral requirement to choose under conditions of uncertainty that is motivated by the kind of considerations that motivate RP cannot permit us to choose against the unanimous preference of the affected parties in a case like that presented in Table 2.

In reply to an earlier draft of this paper, Pettigrew (2023) directly contests this claim. He argues that whereas Buchak’s remarks in justification of RP may support the reading suggested above, there is another way to understand the moral basis of the principle. In his conception, it is not our choice itself, but our choice of risk function that needs to be justifiable to others. Our choice as a whole depends not only on the risk function we use, but also on the probabilities and utilities. The people we affect need not have any right to a say in what probabilities and utilities we use, since these may be settled for us independently by relevant epistemic and moral norms.<sup>7</sup> For example, the way we ought to value different outcomes may be thought of as settled for us by a particular theory of the good and the consequentialist prescription that we morally ought to best promote good consequences. Nonetheless, that prescription fails to determine how we ought to choose in conditions of uncertainty, if many different risk functions are both rationally permissible and reasonable. In that case, none is privileged when it comes to the instrumentally rational pursuit of the good. Since no risk function is so privileged, the decision to elevate a particular risk function as determining our choice needs to be justifiable to those potentially affected. Assuming

<sup>7</sup>Note, however, that Pettigrew (2022a) favours the view that epistemic rationality is permissive and so does not believe in general that probabilities are uniquely determined by relevant epistemic norms.

the probabilities of the different states as objectively determined, we will have done all we need do in terms of justifying our act to the people potentially affected if we can show that our choice maximizes the risk-weighted expectation of aggregate welfare relative to a risk-function derived by aggregating their individual risk attitudes.

This picture is intended as dissolving the apparent tension between RP and consequentialism that I have tried to bring into view, by suggesting that the ideal of respect for others' risk attitudes determines an aspect of moral decision making on which consequentialism itself is silent: namely, which risk function to use in the instrumentally rational pursuit of the good under conditions of uncertainty. However, if we understand consequentialism as the class of moral views according to which the instrumentally rational pursuit of the good is the *only* source of moral reasons, the tension remains. Granting that no risk function from among a given set is privileged when it comes to the instrumentally rational pursuit of the good, the consequentialist concludes that none is privileged *simpliciter*, since only the instrumentally rational pursuit of the good is morally prescribed. A moral requirement to privilege a risk function that is suitably responsive to the risk functions of the people potentially affected by your action therefore must have a non-consequentialist basis. It must be grounded in some kind of ideal of respect for others that is independent of our moral concern to achieve good consequences.

Furthermore, I find it implausible that any such ideal, insofar as it is one we are likely to find compelling, won't also require us to choose Gamble F in Table 2. If my goal is to maximize the welfare of Afryea and Beom-seok and I choose E while knowing that both prefer that I chose F, then insofar as we find it plausible that my decision-making needs to be justifiable to the people potentially affected by my actions in light of an ideal of respect for others that is independent of the aim of achieving morally good consequences, we will surely also be strongly inclined to view my decision as disrespectful to each by virtue of disregarding their preference regarding how best to promote their own welfare. After all, what defence could I give of my choice? Appeal to the fact that choice of E was maximal in respect of the risk-weighted expectation of total welfare relative to their shared risk function seems quite unpersuasive, since the same is true of F and, for that matter, would be true given any risk function, indicating no sensitivity in particular to the risk functions of the people affected by my choice. This suggests that the fact that my choice maximizes the risk-weighted expectation of aggregate welfare relative to the shared risk function of the individuals potentially affected is not in general a sufficient reply to a demand for justification when my choice contravenes the shared preference of those individuals.

Suppose we find my criticisms convincing. Could Pettigrew simply recast his argument, appealing to the aggregate of risk-weighted expected welfare instead as the criterion of choice? Not easily. The most important reason for this is one I'll get to in just a moment. Even apart from the problem I'm going to raise in the next section, we can see that something of the overall tenor of Pettigrew's argument would have to change significantly if recast in these terms. That is because Pettigrew's argument is aimed at total utilitarians and at those who take total utilitarianism to support strong longtermism. However, for the reasons I've highlighted, consequentialists in general cannot adopt preferences over options that track the aggregate of each individual's risk-weighted expected welfare, assuming they accept the weak dominance principle and count outcomes as equally good in which the same number of people are at the same welfare levels, all else being equal.

## 5 | VARIABLE POPULATION CASES

So far, in discussing interpersonal aggregation and respect for others' risk attitudes, I have assumed that the population is fixed. But our ultimate interest is in actions that affect the future

over long-run timescales. In this context, the population is not fixed: the different possible outcomes of at least some acts differ in terms of the number and/or identities of the people who make up the total population of everyone who will ever live. This, I claim, presents a serious challenge to the practical relevance of a principle like RP, insofar as it supports the claim that a gamble is to be morally preferred based on the aggregate of each individual's risk-weighted expected welfare.

Why do I say that? I do so because a person's risk-weighted expected welfare is arguably undefined for any gamble if she does not exist in each possible outcome of the gamble. As Blackorby et al. (2007, 569) note: "individual ex-ante assessments of prospects are meaningless if the person is not alive in all possible states." For any outcome in which she does not exist, a person's welfare level in that outcome will be undefined (see Broome, 1999, 168; Bykvist, 2007; Rabinowicz & Arrhenius, 2015). As a result, her risk-weighted expected welfare is undefined. By extension, the aggregate of each person's risk-weighted expected welfare is undefined.

Of course, the same is true of the aggregate of each person's expected welfare, when we allow 'each person' to range over all those people who may exist as a result of the decision. Suppose, for example, that we aggregate by summing. Then the sum of each person's expected welfare relative to a given act is also undefined when some people exist in only some outcomes and not in others, since every such person's expected welfare is undefined given that their welfare is not well-defined in all possible outcomes. However, given a fixed population and a fixed or shared probability function, the sum of each person's expected welfare is mathematically equivalent to the expectation of the sum of welfare and there is no pressure on us to reason in one way as opposed to the other.<sup>8</sup> In variable population cases, while the sum of each person's expected welfare is undefined, taking into account all those people who may exist when a given act is chosen, it is nonetheless the case that the expectation of total welfare is well-defined, since for each possible outcome, the sum of the welfare of all those individuals existing in that outcome may be assumed to be well-defined. The same is true for the risk-weighted expectation of total welfare. The problem is that choosing on the basis of the risk-weighted expectation of aggregate welfare seems to contravene a plausible ideal of respect for others' risk attitudes of a kind that might motivate RP, for reasons outlined in the previous section.

The conclusion that the aggregate of each person's risk-weighted expected welfare is undefined can, of course, be avoided by ignoring the welfare of all those people who do not exist in every possible outcome of a gamble, treating 'each person' as referring to only those individuals who are guaranteed to exist. We may propose to treat the aggregate of each certainly existing person's risk-weighted expected welfare, as assessed from the perspective of her own risk function (known or imputed), as summarizing one important dimension of what matters morally when choosing in risky contexts. We may, for example, think of this dimension as associated with the Kantian contractualist ideal of justifiability to each person from her own individual perspective, conceding, following Scanlon (1998, 6), that "while it is an important part of morality, as generally understood,

<sup>8</sup> If instead people's subjective probabilities vary, then the aim of maximizing the sum of each individual's expected welfare relative to her own subjective probability function may conflict with the aim of maximising the expectation of the sum of welfare relative to any probability function, for reasons highlighted by Broome (1987). Therefore, even if we insist that prudentially rational agents must be risk neutral with respect to their own welfare, we may come under pressure to choose between acting on the basis of the sum of each person's expectation of their own welfare or the expectation of the sum of welfare in fixed population cases, if we believe agents are morally required to decide in a way that takes account of the different subjective probabilities of the people potentially affected by their actions.

it is only a part, not the whole.” The potential effects of our actions on people who may or may not exist may be thought to belong to a different, more impersonal part of morality.<sup>9</sup>

Indeed, it is well-known that Kantian contractualism has trouble giving an account of variable population cases, even setting aside the problem of risk (see esp. Parfit, 2011, 217–243). In cases that invoke the Non-Identity Problem, if we make the morally worse choice that (certainly) brings about a future population that (certainly) has a significantly lower - but still positive - quality of life than that which would (certainly) have been enjoyed by some other future population we might have created instead, there seems to be no one to whom it can be said that we owed it to them to choose otherwise. Plausibly, we cannot owe it to someone to ensure that they did not come to exist with a life worth living, nor can we owe anything to someone who does not exist. Our reasons for making the morally better choice must therefore be impersonal in nature.<sup>10</sup>

The approach suggested above strikes me as perfectly reasonable, but as rendering an ideal of respect for others' risk attitudes largely irrelevant when thinking about the impact of present actions on the long-run future, by virtue of restricting our attention to potential effects on people who are certain to exist given the choice we make.

A different line of response to my argument that avoids this retreat to near-irrelevance would be to argue that I am mistaken to claim that we cannot decide by appeal to the aggregate of each individual's risk-weighted expected welfare in variable population cases if the term 'each person' is allowed to range over individuals whose existence cannot be guaranteed. We may believe, for example, that a person's welfare level is defined in outcomes in which she does not exist, and that it takes a value of zero (Greaves & Cusbert, 2022; Holtug, 2001; Roberts, 2003). This would indeed undermine my argument. However, it is also a position that I find incredible, since I find it incredible that non-existent people have well-defined levels of any measurable quantity, be it height, weight, or welfare, any more than logic has a well-defined temperature.<sup>11</sup>

Yet another line of response would be to argue that respect for others' risk attitudes does not commit us to choosing in variable population cases by appeal to the aggregate of each individual's risk-weighted expected welfare, as opposed to some other principle that reduces to maximizing the aggregate of each person's risk-weighted expected welfare when dealing with fixed populations, while yielding meaningful and reasonable results in variable population cases. A principle of this kind would appeal to some function of each person's risk-weighted expected welfare conditioning on some event that entails their existence.<sup>12</sup> This might be the event that there exists a certain population of individuals of which they are a member, for example.

However, an immediate problem arises for this proposal, given that we are assuming risk-weighted expected utility theory as our decision theory. Consider some person, Afryea, who may or may not exist given that act  $f$  is chosen. Consider some event  $E$  that may occur if  $f$  is chosen and whose occurrence entails Afryea's existence. We assume that  $E$  does not determine how well Afryea's life in fact goes. Instead, our preference with respect to  $f$  and its alternatives is determined in part by Afryea's risk-weighted expected welfare conditional on  $E$ . We can imagine, for exam-

<sup>9</sup> Kumar (2018, 247) criticizes this proposal on the grounds that “interpersonal obligations are normally understood to be ... a source of decisive reasons” and “take priority over other reasons we have,” whereas what happens to presently existing people should not have absolute priority over what happens to future people. But arguably all this tells us is that our normal understanding of the comparative priority of interpersonal and impersonal obligations is inaccurate.

<sup>10</sup> For proposed contractualist responses to the Non-Identity Problem, see Kumar (2003, 2009, 2018), Finneron-Burns (2016).

<sup>11</sup> This analogy is due to Broome as quoted in Holtug (2001). For discussion, see Holtug (2001, 381–382 n.38).

<sup>12</sup> Compare Harsanyi as quoted in Ng (1983) and Voorhoeve and Fleurbaey (2016, 948–952).

ple, that  $E$  is the existence of a certain population,  $P$ , of which Afryea is a member, and that our preference with respect to  $f$  depends in part on the aggregate of each person in  $P$ 's risk-weighted expected welfare, conditional on  $P$ 's existence.<sup>13</sup> The latter obviously depends in part on Afryea's risk-weighted expected welfare conditional on  $P$ 's existence. Since it is uncertain whether Afryea exists, it is also uncertain whether  $E$  occurs. We need some way to take account of  $E$ 's probability in assessing  $f$ 's choiceworthiness.

The natural way to do this would be to apply risk-weighted expected utility theory. For example, we might imagine associating  $E$  with a value – one that depends in part on Afryea's risk-weighted expected welfare conditional on  $E$  – and then taking the risk-weighted expectation of these values. Thus, we could imagine all the different possible populations, rank them on the basis of the aggregate of each person in the population's risk-weighted expected welfare conditional on the population's existence, and choose based on the risk-weighted expectation of the aggregate values used to rank the possible populations. Call this the *Variable Population Risk Principle* (VPRP). In the fixed population case, VPRP reduces to choosing on the basis of the aggregate of each person's risk-weighted expected welfare. Unlike that criterion, this one can be used in variable population cases without absurdity. Or so it might seem.

The problem is that  $E$  – here, the existence of the population  $P$  – is not a *state of the world* in the technical sense required to apply risk-weighted expected utility theory. A state must uniquely determine an outcome given the decision maker's act, and outcomes must be specified so as to resolve our uncertainty with respect to everything the decision maker cares about.  $E$ 's occurrence does not resolve our uncertainty with respect to everything we care about given choice of  $f$ . In particular, it does not resolve our uncertainty about Afryea's welfare level. Rather than a state of the world,  $E$  is an *event*: i.e., a set of states, and one whose occurrence in conjunction with a given act leaves us uncertain as to the final outcome.

When it comes to the application of expected utility theory, the distinction between events and states is not so important. We can freely swap out a group of states within the expression for the expected utility of an act with an event corresponding to the union of those states, valued at the product of its probability and its expected utility. This doesn't change the expected utility of the act, regardless of how we partition outcomes among events. By contrast, lumping together different possible outcomes into events and calculating risk-weighted expected utility based on the probabilities and risk-weighted utilities of events leads to incompatible valuations of one and the same act depending on how we partition the outcomes (Thoma & Weisberg, 2017). Our calculation of the risk-weighted expected utility of an act therefore has to be based on the utility of outcomes, and cannot be based on the risk-weighted expected utility of events (Buchak, 2013, 226–229). If we rely on calculations that take risk-weighted expectations of the aggregate of each person's risk-weighted expected welfare conditional on uncertain events that entail their existence, we flout this injunction. This is going to get us into trouble.

By way of concrete demonstration, consider the following example (Table 3), which I owe to Timothy Luke Williamson. Consider again VPRP, the view on which we imagine all the different possible populations, rank them on the basis of the aggregate of each person in the population's risk-weighted expected welfare conditional on the population's existence, and choose based on the risk-weighted expectation of the aggregate values used to rank the possible populations. Suppose Afryea is guaranteed to exist, but that it is uncertain whether Beom-seok or Csaba will exist alongside her. Your choice is between gambles G and H. There are three possible states,  $S_1$ ,  $S_2$ ,

<sup>13</sup> A population is here taken to be the collection of every person who exists in a given possible world.

TABLE 3 Mere addition.

	$S_1$	$S_2$	$S_3$
Gamble G	Afryea: 4	Afryea: 10	Afryea: 0
	Beom-seok: $\Omega$	Beom-seok: 0	Beom-seok: 10
	Csaba: 0	Csaba: 0	Csaba: $\Omega$
Gamble H	Afryea: 4	Afryea: 10	Afryea: 0
	Beom-seok: $\Omega$	Beom-seok: 0	Beom-seok: 10
	Csaba: 0	Csaba: 0	Csaba: 1

and  $S_3$ , each with a  $1/3$  probability. In the table below, ' $\Omega$ ' denotes non-existence. As usual, every person has the risk function  $r(\Pr(X)) = \Pr(X)^2$ .

The gambles differ only in that under H, Csaba exists in  $S_3$  and has a life worth living, as opposed to never existing. It would therefore be very strange to have a strict preference for G over H, especially so if that preference is supposed to reflect concern to make our actions justifiable to each person in light of her risk attitude and the uncertainties we faced at the time of decision.<sup>14</sup>

Notice that under G, each state corresponds to a distinct possible population. Given the existence of that population, each person's risk-weighted expected welfare is therefore simply her welfare in the corresponding outcome. Assume that we aggregate by summing. Then taking the risk-weighted expectation of the aggregate of each person's risk-weighted expected welfare within each population as directed by VPRP gives us 6.67.

Under H,  $S_2$  and  $S_3$  yield the same population: namely, the population in which all three people exist. These states are therefore collapsed together into a single event for the purposes of our assessment. Conditional on  $S_2 \cup S_3$ , Afryea's risk-weighted expected welfare is  $10/4$ , Beom-seok's is also  $10/4$ , and Csaba's is  $1/4$ . Summing these, we get  $21/4$ . By contrast, conditional on the existence of the population in state  $S_1$ , each person in the population's risk-weighted expected welfare is just her welfare in the corresponding outcome, and so the sum of risk-weighted expected welfare is 4. Therefore, taking the risk-weighted expectation of the aggregate of each person's risk-weighted expected welfare within each population gives us 4.56.

According to VPRP, then, H is worse than G, although H merely adds an additional person with a life worth living in one of the states. Why does the principle misorder these gambles? For G, each state is handled separately in calculating the risk-weighted expectation of the aggregate of each person's risk-weighted expected welfare. The inputs to the final calculation therefore derive from people's outcome welfare levels, rather than the risk-weighted expectation of their welfare conditional on some event(s). By contrast, in our evaluation of H, VPRP forces us to collapse  $S_2$  and  $S_3$  into a single event whose risk-weighted expected value for each person feeds into the risk-weighted expectation of H as a whole. This leads H to be valued improperly and compared disfavouredly relative to G, although H differs from G only in that Csaba has a life worth living in  $S_3$ .

<sup>14</sup> Someone could conceivably claim that the possible outcomes associated with H are worse in light of Csaba's existence in state  $S_3$ , since Csaba's presence makes the inequality in the corresponding outcome worse. See Parfit (1984, 422–425) for criticism of this sort of view. But even if we are not moved by Parfit's criticisms, it is hard to see a concern to avoid this kind of inequality as reflecting a concern to make our actions justifiable to each possible person, as opposed to reflecting something like the impersonal badness of inequality, as described by Temkin (1993).

## 6 | CONCLUSION

Buchak conjectures that, in combination with RP, any reasonable principle for aggregating across groups of individuals with diverse risk attitudes supports FRAP, the conclusion that decisions that primarily affect future people are required to be governed by a highly risk-avoidant risk attitude. By contrast, I have argued that the considerations that motivate RP justify treating gambles with potential gains and losses for different persons in fixed population cases by appeal to a principle of aggregation that simply breaks down if extended to variable population outcomes of the kind that we inevitably confront when thinking about the future of humanity. There is no principle of interpersonal aggregation of which I know that yields acceptable results in variable population contexts and is consistent with a plausible ideal of respect for others' risk attitudes in fixed population cases.

We should be especially suspicious, I claim, of Pettigrew's argument that RP can be used to derive disturbing conclusions about what actions we should take with respect to the future of humanity given total utilitarianism. That is because the considerations that motivate RP are also of a kind that justify treating gambles with potential gains and losses for different persons in a way that cannot be justified except by departing from the exclusive focus on promoting the good that characterizes consequentialist arguments for strong longtermism of the kind Pettigrew sets out to query. To the extent that respect for others' risk attitudes might incline us to favour voluntary human extinction, that inclination may be one that grips only those of us who reject consequentialism.

## ACKNOWLEDGMENT

For comments on previous drafts of this paper, I'm especially grateful to Lara Buchak and Richard Pettigrew, as well as to Orri Stefánsson, Teru Thomas, Hayden Wilkinson, and Tim L. Williamson, and to audiences at LMU Munich in July 2022, Oxford in August 2022, and at the LSE in May 2023.

## REFERENCES

- Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'école Américaine. *Econometrica*, 21, 503–546.
- Baccelli, J., & Mongin, P. (2021). Can redescription of outcomes salvage the axioms of decision theory? *Philosophical Studies*, 179, 1621–1648.
- Blackorby, C., Bossert, W., & Donaldson, D. (2007). Variable-population extensions of social aggregation theorems. *Social Choice and Welfare*, 28, 567–589.
- Blessenohl, S. (2020). Risk attitudes and social choice. *Ethics*, 130, 485–513.
- Bostrom, N. (2003). Astronomical waste: The opportunity cost of delayed technological development. *Utilitas*, 15, 308–314.
- Bradley, R. (2022). Impartial evaluation under ambiguity. *Ethics*, 132, 541–569.
- Broome, J. (1987). Utilitarianism and expected utility. *Journal of Philosophy*, 84, 405–422.
- Broome, J. (1999). *Ethics Out of Economics*. Cambridge: Cambridge University Press.
- Buchak, L. (2013). *Risk and Rationality*. Oxford: Oxford University Press.
- Buchak, L. (2016). Why high-risk, non-expected-utility-maximising gambles can be rational and beneficial: The case of HIV cure studies. *Journal of Medical Ethics*, 43, 90–95.
- Buchak, L. (2017). Taking risks behind the veil of ignorance. *Ethics*, 127, 610–644.
- Buchak, L. (2019). Weighing the risks of climate change. *The Monist*, 102, 66–83.
- Bykvist, K. (2007). The benefits of coming into existence. *Philosophical Studies*, 135, 335–362.
- Enoch, D. (2016). What's wrong with paternalism: Autonomy, belief, and action. *Proceedings of the Aristotelian Society*, 116, 21–48.
- Finneron-Burns, E. (2016). Contractualism and the non-identity problem. *Ethical Theory and Moral Practice*, 19, 1151–1163.

- Frick, J. (2015). Contractualism and social risk. *Philosophy & Public Affairs*, 43, 175–223.
- Greaves, H., & Cusbert, J. (2022). Comparing existence and non-existence. In J. McMahan, T. Campbell, J. Goodrich, & K. Ramakrishnan (Eds.), *Ethics and Existence: The Legacy of Derek Parfit* (pp. 61–92). Oxford: Oxford University Press.
- Greaves, H., & MacAskill, W. (2021). *The case for strong longtermism*. (Global Priorities Institute Working Paper No. 5-2021. <https://globalprioritiesinstitute.org/wp-content/uploads/The-Case-for-Strong-Longtermism-GPI-Working-Paper-June-2021-2-2.pdf>. Last accessed 11 December 2023).
- Groll, D. (2012). Paternalism, respect and the will. *Ethics*, 122, 692–720.
- Holtug, N. (2001). On the value of coming into existence. *The Journal of Ethics*, 5, 361–384.
- Kelly, T. (2013). Evidence can be permissive. In M. Steup & J. Turri (Eds.), *Contemporary Debates in Epistemology* (pp. 298–311). Oxford: Blackwell.
- Kumar, R. (2003). Who can be wronged? *Philosophy & Public Affairs*, 31, 99–118.
- Kumar, R. (2009). Wronging future people. In A. Gosseries & L. Meyer (Eds.), *Intergenerational Justice* (pp. 251–273). Oxford: Oxford University Press.
- Kumar, R. (2018). Risking future generations. *Ethical Theory and Moral Practice*, 21, 245–257.
- Nebel, J. M. (2021). Rank-weighted utilitarianism and the veil of ignorance. *Ethics*, 131, 87–106.
- Ng, Y.-K. (1983). Some broader issues of social choice. In P. K. Pattanaik & M. Salles (Eds.), *Social Choice and Welfare* (pp. 151–173). Amsterdam: North-Holland.
- Parfit, D. (1984). *Reasons and Persons*. Oxford: Oxford University Press.
- Parfit, D. (2011). *On What Matters, Volume 2*. Oxford: Oxford University Press.
- Pettigrew, R. (2022a). *Epistemic Risk and the Demands of Rationality*. Oxford: Oxford University Press.
- Pettigrew, R. (2022b). *Should longtermists recommend hastening extinction rather than delaying it?* (Global Priorities Institute Working Paper No. 2-2022. <https://globalprioritiesinstitute.org/wp-content/uploads/Pettigrew-Effective-altruism-risk-and-human-extinction-2.pdf>. Last accessed 11 December 2023).
- Pettigrew, R. (2023). *Utilitarianism and risk: A reply to Mogensen*. <https://m-phi.blogspot.com/2023/06/utilitarianism-and-risk-reply-to.html>. Last accessed 11 December 2023.
- Quong, J. (2010). *Liberalism Without Perfection*. Oxford: Oxford University Press.
- Rabinowicz, W. & Arrhenius, G. (2015). The value of existence. In I. Hirose & J. Olson (Eds.), *The Oxford Handbook of Value Theory* (pp. 424–444). Oxford: Oxford University Press.
- Raiffa, H. (1968). *Decision Analysis: Introductory Lectures on Choices Under Uncertainty*. New York, NY: Random House.
- Rawls, J. (1993). *Political Liberalism*. New York, NY: Columbia University Press.
- Roberts, M. A. (2003). Can it ever be better never to have existed at all? Person-based consequentialism and a new repugnant conclusion. *Journal of Applied Philosophy*, 20, 159–185.
- Scanlon, T. M. (1998). *What We Owe To Each Other*. Cambridge, MA: Harvard University Press.
- Schoenfield, M. (2014). Permission to believe: Why permissivism is true and what it tells us about irrelevant influences on belief. *Noûs*, 48, 193–218.
- Shiffrin, S. V. (2000). Paternalism, unconscionability doctrine, and accommodation. *Philosophy and Public Affairs*, 29, 205–250.
- Slovic, P., & Tversky, A. (1974). Who accepts Savage's axiom? *Behavioral Science*, 19, 368–373.
- Steel, D. (2015). *Philosophy and the Precautionary Principle: Science, Evidence, and Environmental Policy*. Cambridge: Cambridge University Press.
- Stefánsson, H. O., & Bradley, R. (2019). What is risk aversion? *The British Journal for the Philosophy of Science*, 70, 77–102.
- Temkin, L. S. (1993). *Inequality*. Oxford: Oxford University Press.
- Thoma, J. (2023). Taking risks on behalf of another. *Philosophy Compass*, 18, e12898.
- Thoma, J., & Weisberg, J. (2017). Risk writ large. *Philosophical Studies*, 174, 2369–2384.
- Voorhoeve, A., & Fleurbaey, M. (2016). Priority or equality for possible people? *Ethics*, 126, 929–954.

**How to cite this article:** Mogensen, A. L. (2024). Respect for others' risk attitudes and the long-run future. *Noûs*, 58, 1017–1031. <https://doi.org/10.1111/nous.12488>