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

Andreas Mogensen (Global Priorities Institute, University of Oxford)

Global Priorities Institute | December 2022

GPI Working Paper No . 20-2022

GLOBAL PRIORITIES INSTITUTE

UNIVERSITY OF OXFORD

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

Andreas L. Mogensen

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

The long-run future is highly uncertain, as are the effects of present actions on posterity. In order to be able to make reasonable decisions that take account of the potential long-term impact of our choices today, we therefore need to know how to rationally manage uncertainty in decision-making.

According to orthodox decision theory, a rational agent in conditions of uncertainty prefers those acts that maximize expected utility (Arnauld and Nicole 1662; Bernoulli 1738; Ramsey 1926; von Neumann and Morgenstern 1947; Savage 1972). The utility function is assumed to be a cardinal measure of the agent's strength of preferences over outcomes,

and its expectation is taken relative to a probability function representing known chances and/or the strength of the agent's beliefs about the state of the world.

In the recent philosophical literature, an influential alternative to expected utility theory is defended by Buchak (2013), building on earlier work by Quiggin (1982). Buchak 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 states of the world and the desirability 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 moral contexts require that we adopt a particular attitude toward risk. We are required, she claims, to 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. This approach has also been claimed to have important consequences for evaluating the prospect of continued human survival and actions aimed at ensuring that our species endures. In particular, Pettigrew (2022) argues that consequentialists are pushed 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 this line 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 the risk attitudes of the people we potentially affect. So goes the thought. I will offer reasons to think that a plausible ideal of respect for others' risk attitudes does not support the kind of conclusions outlined above, and may even be irrelevant when thinking about the impact of present actions on the long-run future. The problem is that there is no known principle of interpersonal aggregation suitable to variable population contexts that is consistent with a plausible ideal of respect for others' risk attitudes in fixed population cases.

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 an important choice that we face when we aim to respect others' risk attitudes when choosing on behalf of a group of persons: namely, whether to aggregate first across persons and then across outcomes, or *vice versa*. In section 5, 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 6, I argue that this approach also threatens to break down in variable population cases of the kind we inevitably confront in making decisions about the long-run future. The same is not true of a procedure that aggregates across persons within outcomes and then across outcomes, but this procedure cannot be justified by appeal to a plausible ideal of respect for others' risk attitudes. Section 7 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 that an agent has available different possible acts, and that, together with the state of the world, which act she chooses will determine what outcome she brings about. Thus, assume there are n possible states of the world, $S_1, \ldots, S_n$. If act f is chosen, corresponding outcomes $x_1, \ldots, x_n$ will be realized depending on which of $S_1, \ldots, S_n$ is actual. We assume the agent's beliefs about which state is actual are represented by a subjective probability function, $\Pr(\cdot)$. Furthermore, we assume that the agent's preferences over outcomes are represented by a cardinal utility function, $u(\cdot)$. Lastly, we assume that states and outcomes are indexed in ascending preference order, such that $u(x_i) \leq u(x_{i+1})$ for $i = 1, \ldots, n-1$. 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.

Note that risk avoidance, so understood, is conceptually distinct from *risk aversion* (see Buchak 2013: 62–66). As just defined, an agent is risk avoidant just in case her risk function is strictly convex. An agent is said to be risk averse in respect of some good just in case she always prefers a given uncertain prospect involving different possible amounts of that good to any *mean-preserving spread*. Roughly speaking, a mean-preserving spread of a prospect adds to each possible outcome of that prospect a randomly determined gain or loss, whose expected value is zero conditional on any outcome. According to orthodox decision theory, an agent can be risk averse with respect to some good only by virtue of valuing outcomes involving different amounts of that good in a certain way: risk aversion in respect of some good reduces to having a utility function that is strictly concave in that good; equivalently, the good must exhibit diminishing marginal utility for the agent. However, agents who are risk avoidant in Buchak's sense can be risk averse in respect of some good even if their utility function is linear in that good, with every increasing increment of the good valued the same amount.[1]

REU theory can explain some cases of apparently rational decision-making under con-

[1] See also Stefánsson and Bradley (2019) for a framework that allows for risk aversion without diminishing marginal utility for outcomes, based on Jeffrey-Bolker decision theory (Jeffrey 1965).

ditions of risk that cannot be captured by expected utility theory. Consider the famous *Allais paradox* (Allais 1953), to which we'll return later on. Imagine a lottery in which there are 100 tickets and four possible gambles grouped into two pairs, yielding different payoffs depending on which ticket is drawn:

	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

Table 1: The Allais Paradox

Thus, A gives you an 89% chance of \$1 million, a 10% chance of \$5 million, and a 1% chance of nothing, whereas B guarantees you \$1 million. C gives you a 10% chance of \$5 million and a 90% chance of nothing, whereas D gives you an 11% chance of \$1 million and an 89% chance of nothing. In choices of this kind, people typically prefer B to A but also prefer C to D (Kahneman and Tversky 1979). These preferences seem inconsistent in principle with valuing an uncertain prospect at its expected utility,[2] but can be accommodated by the assumption that agents maximize risk-weighted expected utility and are risk avoidant.[3]

[2]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.

[3]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 otherwise agree in their moral outlook and their empirical beliefs. For example, when it comes to the evaluation of continued human survival and the available acts aimed at that goal, a group of total utilitarians whose utility functions are linear functions of total welfare and who assign the same probabilities to the possible states of the world may nonetheless permissibly come to very different conclusions by virtue of assessing the value of continued human survival in terms of the risk-weighted expectation of total welfare using their own idiosyncratic risk function, which may be linear, strictly convex, or strictly concave *ad libitum*.

However, there is nothing obvious about this conclusion, insofar as we are talking about moral permissibility. After all, just because rationality *per se* does not constrain an agent's utility function in some way does not entail that morality is similarly permissive. It may not be irrational to be indifferent to the suffering of innocents, but it is unethical. It might similarly be true that there are many rationally permissible risk functions that it would be morally wrong to choose in light of in a given context.

Buchak (2019) adopts a view along exactly these lines. 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: 78) Morally speaking, the agent's own idiosyncratic risk function must be set aside – unless it happens to be very risk avoidant.

How does Buchak arrive at FRAP? To begin, consider the general question of whether our evaluation of gambles over other people's welfare should defer to the risk attitudes of the people potentially affected by our actions (Thoma 2021). When our actions affect other people, it seems plausible that *our* preferences over outcomes morally ought to be informed by *their* self-regarding preferences over outcomes. When the outcome associated with any given act is unknown, should the risk function that we use in making our

decision also be informed by the risk attitudes of the people potentially affected by our actions? Or can we go ahead and happily use our own, potentially idiosyncratic risk function?

According to Buchak, 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,[4] so that our choice cannot reasonably be rejected as excessively risky by those we might affect.[5]

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) By way of intuitive support for the first half of RP, Buchak asks us to imagine that an acquaintance has an injury that may be either a spasm, which can be completely relieved with the application of heat, or a pulled muscle, which can be partially treated to the point of yielding only mild pain with the application of ice. The catch is that heat yields intense pain if applied to a pulled muscle, whereas ice elicits only moderate pain when applied to a spasm. Intuitively, Buchak claims, if we do not know this person's own attitude toward risk, we ought to apply ice, rather than subject

[4]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.

[5]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.

our acquaintance to the riskier prospect associated with the application of heat, unless the injury is very unlikely to be a pulled muscle. This intuition is taken to support RP.

Buchak then arrives at FRAP via RP. Very roughly, her argument is 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, 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), placing significantly greater relative weight on the avoidance of worst-case outcomes than would be recommended by a policy of maximizing expected value.

More recently, Pettigrew (2022) 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 that total utilitarianism supports the disturbing conclusion that we are morally required 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 and MacAskill, 2021: 3).

4 Interpersonal Aggregation

Of itself, RP governs only the case of 'making a decision for an individual'. What about the case of deciding for a group 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?

Let's assume for the time being that the population is fixed. I will also assume that moral preferences over outcomes are based on aggregate welfare. The phrase 'aggregate welfare' has a utilitarian air to it, so I want to emphasize that my assumption does not commit us to a utilitarian moral theory, nor a utilitarian axiology. It is intended to be compatible with any way of aggregating welfare across persons, including, say, non-Archimedean views on which comparatively trivial benefits and losses are to be ignored entirely when we aggregate (Voorhoeve 2014).

Here, then, is the choice we face (see Blessenohl 2020). When choosing on behalf of a group of persons in a way that is respectful of others' risk attitudes, one possibility 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. Another possibility is 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.[6]

To see the contrast, it helps to work through an example. In order to do so, it will be helpful, for the sake of simplicity, to temporarily adopt a utilitarian ranking of outcomes that just sums individual welfare levels. To avoid any complications about how to aggregate people's risk functions so as to arrive at a collective risk function, it is also helpful to assume that people all have or are imputed to have the same risk function. I assume that in this case, the collective risk function should be the consensus risk function.

[6] The choice we face here is similar in important respects to the choice between *ex post* and *ex ante* egalitarianism (Segall 2016: 205–224; Voorhoeve and Fleurbaey 2012, 2016), *ex post* and *ex ante* prioritarianism (Adler 2012: 477–451; Holtug 2018; McCarthy 2006, 2008; Rabinowicz 2002; Voorhoeve and Fleurbaey 2016), and *ex post* and *ex ante* contractualism (Frick 2015; Kumar 2005; Otsuka 2015; Reibetanz 1998).

Thus, suppose that we must choose between a status quo outcome in which Afryea and Beom-seok each have welfare 5 and a coin flip that with .5 probability gives Afryea welfare 11 and Beom-seok welfare 0 and with .5 probability instead gives Afryea welfare 0 and Beom-seok welfare 11. Suppose it is known that both Afryea and Beom-seok have the risk function $r(\Pr(X)) = \Pr(X)^2$.

	The coin lands Heads	The coin lands Tails
Status Quo	Afryea: 5 Beom-seok: 5	Afryea: 5 Beom-seok: 5
Coin Flip	Afryea: 11 Beom-seok: 0	Afryea: 0 Beom-seok: 11

Table 2: Status Quo vs Coin Flip

Notice that there is no gamble over total welfare, only with respect to each person's share of it. For each option, the outcome is certain if described in units of total welfare: either 10 given the status quo or 11 given the coin flip. Therefore, if we choose on the basis of the risk-weighted expectation of total welfare, we choose simply on the basis of total welfare, and so we prefer to flip the coin.

Suppose instead that we choose based on the sum of risk-weighted expected welfare. In the status quo option, there is as yet no uncertainty. Each person's welfare is guaranteed to be 5, and so is their risk-weighted expected welfare. Therefore, the sum of each person's risk-weighted expected welfare is $2 \times 5 = 10$. Under the coin flip, however, each person's risk-weighted expected welfare is $[0 + (0.5^2 \times (11 - 0))] = 2.75$, so the sum of the risk-weighted expectation of each person's welfare is $2 \times 2.75 = 5.5$. Therefore, if we choose on the basis of the sum of each person's risk-weighted expected welfare, we prefer the status quo.

5 Respect For Others' Risk Attitudes and (Non-) Consequentialism

Besides the fact that the different approaches to aggregation yield different results, I also want to emphasize the following point about Table 2. For reasons I have already noted, if we accept a utilitarian axiology, then the outcome is guaranteed to be better if we flip the coin, no matter how the coin falls, since total welfare is guaranteed to increase from 10 to 11. In other words, the coin flip dominates the status quo, assuming a utilitarian axiology. Therefore, if we choose based on the sum of each individual's risk-weighted expected welfare and so prefer the status quo, that must be because we are morally concerned with something other than the value of outcomes, conceived as total welfare.

This line of argument can be pushed further, suggesting that 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 agent-neutral value of outcomes is the only determinant of what we morally ought to choose. More exactly, we assume that consequentialism entails a *weak dominance principle*, on which we morally ought not to strictly prefer one gamble to another if the latter yields an outcome that is morally at least as good in every possible state of the world. Then the incompatibility of consequentialism and choosing on the basis of the aggregate of each individual's risk-weighted expected welfare can be generalized to any welfarist axiology that obeys a *principle of anonymity*, according to which outcomes are equally good in which the same number of people are at the same welfare levels, and any principle for aggregating risk-weighted expected welfare that obeys an *ex ante Pareto condition*, such that if one gamble has greater risk-weighted expected welfare for each person, then it has greater aggregate risk-weighted expected welfare (Blessenohl 2020; Nebel 2021).

To push the argument further in this way, we will need a different example. After all, we might have a different conception of the value of outcomes from that adopted by utilitarians, and if we reject utilitarianism in favour of, say, prioritarianism, then we may well deny that the coin flip is guaranteed to yield a better outcome in the example represented

in Table 2.[7]

Consider, then, the following example, based on Nebel (2021: 103–104). Imagine that both Afryea and Beom-seok have zero wealth at present and have welfare functions that are the square root of their dollar holdings denominated in millions. Both have the risk funtion $r(\Pr(X)) = \Pr(X)^2$ and so in maximizing their risk-weighted expected welfare, each exhibits the Allais preferences: each prefers B to A and C to D. Imagine that we have the following choice:

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

Table 3: The Allais Paradox Rides Again

Thus, E gives Afryea an 89% of \$1M, 10% chance of \$5M, and 1% chance of nothing, whereas F guarantees her \$1M, and E gives Beom-seok 11% chance of \$1M and 89% chance of nothing, whereas F gives Beom-seok a 10% chance of \$5M and a 90% chance of nothing. 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.

By stipulation, Afryea has greater risk-weighted expected welfare under B than A and Beom-seok has greater risk-weighted expected welfare under C than D, so each has greater risk-weighted expected welfare under Gamble F. By the *ex ante* Pareto condition, F is preferred. However, any axiology that obeys the anonymity principle entails that the out-

[7] Notably, Buchak (2017) argues that RP supports a form of prioritarianism, given an Ideal Contractarian theory of justice that identifies justice with what the parties would choose from behind the Veil of Ignorance (Vickrey 1945, Harsanyi 1953, Rawls 1971).

comes associated with E and F are equally morally good no matter which arises. Any principle of choice under uncertainty that is guided by aggregating each person's risk-weighted expected welfare therefore must be morally concerned with something other than the value of outcomes, assuming that the aggregation function satisfies the *ex ante* Pareto condition, the value function on outcomes satisfies anonymity, and any consequentialist choice criterion satisfies the weak dominance principle.

What does RP prescribe in cases like this? Well, RP is, strictly speaking, silent about cases of this sort. As I have noted, it is formulated in such a way that it only applies to the case of 'choosing for an individual'. Nonetheless, the considerations that motivate RP seem to strongly favour choice of F over E.

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 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 cannot be made consistent with choosing among options based on exclusive concern for achieving morally good outcomes. In a sense, this should be unsurprising. 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). That theory is obviously designed to be sensitive to moral considerations that are not grounded in facts about the moral goodness of outcome. 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 con-

cerned with the value of outcomes for them, as on Buchak's theory (Thoma 2021). It is much easier to motivate on a conception of the moral domain on which respect for the autonomous personhood of others and a desire to find principles that we can all collectively will and endorse provide the foundations of right and wrong. In other words, RP has a natural place within a broadly Kantian moral framework, but not within a consequentialist moral theory.

The points I've just gone over are especially important for assessing Pettigrew's argument. On the one hand, Pettigrew's approach to choosing on behalf of groups of individuals is the contrary of the approach I have claimed is recommended to us insofar as we find RP plausible. Pettigrew (2022: 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 $r_1, \ldots, r_n$, Pettigrew suggests that the collective risk function r_G might be obtained by taking a weighted average, $r_G = \lambda_1 r_1 + \cdots + \lambda_n r_n$, where the weights sum to 1 and λ_i is greater the more risk avoidant r_i 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 are difficult to square with a plausible ideal of respect for others' risk attitudes.

Pettigrew (2022: 32–4) does address the objection that his preferred decision procedure requires 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 will require us to prefer the option that violates the unanimous self-regarding preferences of the affected parties in a case of this kind when outcomes are valued at their total welfare and we aim to maximize the good, given that choice of Coin Flip in Table 2 is guaranteed to yield a better outcome if outcomes are valued thusly. Even when individuals' self-regarding preferences over gambles are based on their expected welfare, the goal of maximizing total welfare can in some cases 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; Schoenfield 2014) (compare Blessenohl 2020: 499–501).

However, the objection I am raising is not that going against the unanimous preference of the affected parties in a case like that presented in Table 2 gives the wrong result. To raise that objection would be to presupposes the falsity of total utilitarianism considered as a theory of right action, whereas my objection is intended as neutral with respect to the correct moral theory. My objection is that a moral requirement to choose under conditions of uncertainty that is motivated by the kind of considerations that motivate RP cannot require us to choose against the the unanimous preference of the affected parties in a case like that presented in Table 2. An approach that delivers that result cannot be justified in terms of a plausible ideal of respect for others' risk-attitudes.

Could Pettigrew recast his argument, appealing to the aggregate of risk-weighted expected welfare 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.

Recall that Pettigrew's argument is aimed at total utilitarians and at those who take total utilitarianism to support strong longtermism. For the reasons I've highlighted, total utilitarians cannot accept that we should choose on behalf of a group of individuals by reference to the aggregate of each individual's risk-weighted expected welfare, since doing so sometimes recommends choice of an option that is guaranteed to be worse in respect of total welfare. 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 the anonymity principle on which outcomes are equally good in which the same number of people are at the same welfare levels, all else being equal.

6 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: 16; Bykvist 2007; Rabinowicz and Arrhenius 2015). As a result, both her expected welfare and her risk-weighted expected welfare are undefined. By extension, the aggregate of each person's risk-weighted expected welfare is undefined.

This conclusion 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' in 'the aggregate of each person's risk-weighted expected welfare' 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 [this] is an important part of morality, as generally understood, 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. [8]

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.[9]

The approach suggested above strikes me as acceptable in principle, 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 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 takes a value of zero (Holtug 2001, Roberts 2003, Greaves and Cusbert 2022). 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 lev-

[8] Kumar (2018: 247) criticizes this proposal on the grounds that "intepersonal 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.

[9] For proposed contractualist responses to the Non-Identity Problem, see Kumar (2003, 2009, 2018), Finneron-Burns (2016).

els of any measurable quantity, be it height, weight, or welfare, any more than logic has a temperature.[10]

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.[11] 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 example, 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.[12] 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

[10] This analogy is due to Broome as quoted in Holtug (2001). For discussion, see Holtug (2001: 381–382 n.38).

[11] Compare Harsanyi as quoted in Ng (1983) and Voorhoeve and Fleurbaey (2016: 948–952).

[12] A population is here taken to be the collection of every person who exists in a given possible world.

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

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

Table 4: Mere Addition

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.[13]

[13]Someone could conceivably claim that the possible outcomes associated with H are worse in light of Csaba's

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 4.55.

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 disfavourably relative to G, although H differs from G only in that Csaba has a life worth living in S_3.

---

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).

7 Conclusion

Recall again Buchak's 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" (Buchak 2019: 78). As noted previously, Buchak justifies FRAP as follows. Since the risk attitudes of future people are unknown to us, RP is taken to support the conclusion that 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, we should expect that decisions that primarily affect future people are required to be governed by a highly risk avoidant risk attitude.

I have argued that the final step in this argument is false. In particular, the considerations that motivate RP are of a kind that justify treating gambles with potential gains and losses for different persons in fixed population cases by appeal to a principle of aggregation that 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.

References

Adler, M. 2012. *Well-Being and Fair Distribution: Beyond Cost-Benefit Analysis*. Oxford: Oxford University Press.

Allais, M. 1953. Le comportement de l'homme rationnel devant le risque: Critique des postulats et axioms de l'ecole Americaine. *Econometrica 21*(4): 503–546.

Arnauld, P. and P. Nicole. 1662. *La Logique, ou l'Art de Penser*. Paris: Jean Guignart, Charles Savreux, Jean de Lavnay.

Baccelli, J. and P. Mongin. 2021. Can redescriptions of outcomes salvage the axioms of decision theory? *Philosophical Studies 179*(5): 1621–1648.

Bernoulli, D. 1738. Specimen theoriae novae de mensura sortis. *Papers of the Imperial Academy of Sciences in Petersburg* 5: 175–192.

Blackorby, C., W. Bossert, and D. Donaldson. 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*(4): 485–513.

Bostrom, N. 2003. Astronomical waste: The opportunity cost of delayed technological development. *Utilitas 15*(3): 308–314.

Broome, J. 1987. Utilitarianism and expected utility. *Journal of Philosophy 84*(8): 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*(2): 90–95.

Buchak, L. 2017. Taking risks behind the veil of ignorance. *Ethics 127*(3): 610–644.

Buchak, L. 2019. Weighing the risks of climate change. *The Monist 102*(1): 66–83.

Bykvist, K. 2007. The benefits of coming into existence. *Philosophical Studies 135*(3): 335–362.

Finneron-Burns, E. 2016. Contractualism and the non-identity problem. *Ethical Theory and Moral Practice 19*(5): 1151–1163.

Frick, J. 2015. Contractualism and social risk. *Philosophy & Public Affairs 43*(3): 175–223.

Greaves, H. and J. Cusbert. 2022. Comparing existence and non-existence. In *Ethics and Existence: The Legacy of Derek Parfit*, eds. J. McMahan, T. Campbell, J. Goodrich, and K. Ramakrishnan, 61–92. Oxford: Oxford University Press.

Greaves, H. and W. MacAskill. 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.

Harsanyi, J. C. 1953. Cardinal utility in welfare economics and in the theory of risk-taking. *The Journal of Political Economy 61*(5): 434–435.

Holtug, N. 2001. On the value of coming into existence. *The Journal of Ethics 5*(4): 361–384.

Holtug, N. 2018. Prioritarianism: Ex ante, ex post, or factualist criterion of rightness? *The Journal of Political Philosophy 27*(2): 207–228.

Jeffrey, R. C. 1965. *The Logic of Decision*. New York, NY: University of Chicago Press.

Kahneman, D. and A. Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica 47*(2): 263–292.

Kelly, T. 2013. Evidence can be permissive. In *Contemporary Debates in Epistemology*, eds. M. Steup and J. Turri, 298. Oxford: Blackwell.

Kumar, R. 2003. Who can be wronged? *Philosophy & Public Affairs 31*(2): 99–118.

Kumar, R. 2005. Risking and wronging. *Philosophy & Public Affairs 43*(1): 27–51.

Kumar, R. 2009. Wronging future people. In *Intergenerational Justice*, eds. A. Gosseries and L. Meyer, 251–273. Oxford: Oxford University Press.

Kumar, R. 2018. Risking future generations. *Ethical Theory and Moral Practice 21*(2): 245–257.

McCarthy, D. 2006. Utilitarianism and prioritarianism I. *Economics & Philosophy 22*(3): 335–363.

McCarthy, D. 2008. Utilitarianism and prioritarianism II. *Economics & Philosophy 24*(1): 1–33.

Nebel, J. M. 2021. Rank-weighted utilitarianism and the veil of ignorance. *Ethics 131*(1): 87–106.

Ng, Y.-K. 1983. Some broader issues of social choice. In *Social Choice and Welfare*, eds. P. K. Pattanaik and M. Salles, 151–173. Amsterdam: North-Holland.

Otsuka, M. 2015. Risking life and limb: How to discount harms by their improbability. In *Identified Versus Statistical Lives: An Interdisciplinary Perspective*, eds. I. G. Cohen, N. Daniels, and N. Eyal, 77–93. Oxford: Oxford University Press.

Parfit, D. 1984. *Reasons and Persons.* Oxford: Oxford University Press.

Parfit, D. 2011. *On What Matters, Volume 2.* Oxford: Oxford University Press.

Pettigrew, R. 2022. 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.

Quiggin, J. 1982. A theory of anticipated utility. *Journal of Economic Behavior Organization 3*(4): 323–343.

Rabinowicz, W. 2002. Prioritarianism for prospects. *Utilitas 14*(1): 2–21.

Rabinowicz, W. and G. Arrhenius. 2015. The value of existence. In *The Oxford Handbook of Value Theory*, eds. I. Hirose and J. Olson, 424–444. Oxford: Oxford University Press.

Raiffa, H. 1968. *Decision Analysis: Introductory Lectures on Choices Under Uncertainty.* New York, NY: Random House.

Ramsey, F. P. 1926. Truth and probability. In *Foundations of Mathematics and Other Logical Essays*, 156–198. Abingdon: Routledge.

Rawls, J. 1971. *A Theory of Justice*. Cambridge, MA: Harvard University Press.

Rawls, J. 1993. *Political Liberalism*. New York, NY: Columbia University Press.

Reibetanz, S. 1998. Contractualism and aggregation. *Ethics 108*(2): 296–311.

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*(2): 159–185.

Savage, L. J. 1972. *Foundations of Statistics, Second Edition*. New York, NY: Dover.

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*(2): 193–218.

Segall, S. 2016. *Why Inequality Matters: Luck Egalitarianism, its Meaning and Value*. Cambridge: Cambridge University Press.

Steel, D. 2015. *Philosophy and the Precautionary Principle: Science, Evidence, and Environmental Policy*. Cambridge: Cambridge University Press.

Stefánsson, H. O. and R. Bradley. 2019. What is risk aversion? *The British Journal for the Philosophy of Science* 70.

Temkin, L. S. 1993. *Inequality*. Oxford: Oxford University Press.

Thoma, J. 2021. Taking risks on behalf of another. https://johannathoma.files.wordpress.com/2021/08/taking-risks-on-behalf-of-another.pdf.

Thoma, J. and J. Weisberg. 2017. Risk writ large. *Philosophical Studies 174*(9): 2369–2384.

Vickrey, W. 1945. Measuring marginal utility by reactions to risk. *Econometrica 13*(4): 319–333.

von Neumann, J. and O. Morgenstern. 1947. *Theory of Games and Economic Behavior, Second Edition*. Princeton, NJ: Princeton University Press.

Voorhoeve, A. 2014. How should we aggregate competing claims? *Ethics 125*(1): 64–87.

Voorhoeve, A. and M. Fleurbaey. 2012. Egalitarianism and the separateness of persons. *Utilitas 24*(3): 381–398.

Voorhoeve, A. and M. Fleurbaey. 2016. Priority or equality for possible people? *Ethics 126*(4): 292–254.