

Taking Risks on Behalf of Another

August 29, 2021

Abstract

A growing number of decision theorists have, in recent years, defended the view that rationality is permissive in the sense that there is rational leeway in how agents who value outcomes in the same way may choose under risk, allowing for different levels of risk aversion or risk inclination. One new question that arises once we grant such permissiveness is what attitude to risk we should implement when choosing on behalf of another person. Are we permitted to implement any of the rationally permissible risk attitudes, is there some specific risk attitude that is required when choosing for others, or are we required to defer to the risk attitudes of the person on whose behalf we are choosing? This article elaborates on this question, explains its wider practical and theoretical significance, provides an overview of existing answers, and explores how to go about providing a more systematic account of how to choose on behalf of others in risky contexts.

1 Introduction: Permissiveness of Rationality under Risk

A growing number of decision theorists have, in recent years, defended the view that rationality is permissive under risk in the following sense. Suppose an agent does not know what states of affairs her choices will ultimately bring about, but she values those potential states of affairs – call them outcomes – to different extents. Under risk, agents can assign probabilities to how likely different outcomes are to come about as a consequence of the available options. Agents thus choose between outcome-gambles. If rationality is permissive under risk, probabilities together with an agent’s valuations of outcomes do not fully determine how she should rationally choose under risk; two agents who share the same probability assignments and value outcomes in exactly the same way may nevertheless make different choices between risky gambles. In particular, rational agents may, within limits, exhibit more or less risk averse preferences and choice behaviours in their pursuit of valuable outcomes.

The following example illustrates this idea. Suppose you have to choose between two different medical treatments to undergo. When evaluating the outcomes these treatments may bring about, all you care about is some measure of quality-adjusted life-years

(QALYs), suitably tailored to your desires and circumstances. Moreover, QALYs are a cardinal measure of your values: Losing two QALYs is twice as bad as losing one, and so on.¹ On the best available evidence, the two treatments available to you have the following prospects:

Treatment 1: Lose 2 QALYs for certain.

Treatment 2: 10% chance of losing 18 QALYs (and none otherwise).

Which treatment should you choose? A risk neutral agent would maximise the expectation of QALYs, and go for *Treatment 2*. But *Treatment 1* may seem attractive to you if you would rather not risk the terrible outcome of losing 18 QALYs. *Treatment 1* is a more risk averse course of action. Intuitively, both the risk neutral and the risk averse choice seem rationally permissible. More generally, within limits (for instance, ruling out extreme risk-seeking or risk aversion, and violation of plausible axioms such as state-wise dominance) there seems to be rational leeway allowing different risk attitudes in our pursuit of good outcomes. Those who defend the permissiveness of rationality under risk aim to accommodate this intuition.

Permissiveness under risk can be accommodated in a variety of different ways in formal decision theories. Let me briefly explain three approaches here, though my discussion below aims to apply equally to each. The first approach is to accommodate it within orthodox expected utility theory in the von Neumann and Morgenstern (1944) and Savage (1972) traditions. Next to probabilities, orthodox decision theory features utilities, which are assigned to outcomes, and only to outcomes. As the name suggests, the theory then requires agents to maximise the expectation, that is, the probability-weighted sum of outcome-utilities. Now we want to accommodate risk averse (or risk-seeking) preferences that don't maximise the expectation of the true cardinal measure of an agent's values (QALYs in our case). To do so within orthodox expected utility theory, we must give up on the standard interpretation of utility as a cardinal measure of an agent's valuation of outcomes. Utility and QALYs can't be equivalent in our case, despite QALYs being a cardinal measure of everything the agent values in an outcome. Indeed, the plausibility of the kind of risk aversion just described is sometimes appealed to as an argument in favour of giving up such a 'realist' interpretation of utility in expected utility theory, and in favour of thinking about utility as a construct used to conveniently represent preferences but with a less definite psychological and normative interpretation (see Broome 1991 and Velleman 1993). To accommodate risk aversion or risk-seeking regarding value itself within orthodox expected utility theory, the shape of the utility function must in part capture the agent's risk attitudes, on top of capturing her valuation of outcomes. For instance, for risk

¹The ambition for QALY measures in use in health policy evaluation is indeed that they provide such a cardinal measure: The quality-adjustment is meant to measure the desirability of time spent in a certain health state, and agents are assumed to value time spent in any particular health state linearly in time. Note that some prominent methods for obtaining such a cardinal measure in fact assume a risk neutral version of expected utility theory (e.g. standard-gamble methods), and thus rule out permissiveness under risk as we understand it here. However, others, such as the use of visual analogue scales, do not. See Weinstein et al. (2009) for an introductory overview of QALY measurement.

averse agents, the marginal utility of a true cardinal measure of value would be decreasing. Permissiveness of rationality under risk finds expression in permissiveness about the shape of the utility function holding fixed valuation of outcomes.

A second way of accommodating permissiveness under risk involves adopting alternative decision theories that explicitly represent attitudes to risk separately from utility and probability. A non-expected utility theory of much recent popularity is Buchak’s (2013) risk-weighted expected utility theory, based on the rank-dependent utility theory due to Quiggin (1982). In this decision theory, the utilities of outcomes are not only weighted by their probabilities, but also using a risk function to arrive at an overall evaluation. This allows agents to give greater or smaller weight to especially bad potential outcomes, capturing greater or smaller levels of risk aversion. When risk attitudes are captured by the risk function, utility can again function as a cardinal measure of the value of outcomes. In this approach, permissiveness of rationality under risk amounts to permissiveness about the risk function.

A final way of accommodating permissiveness under risk is to allow for utilities to be assigned to objects other than outcomes, in a way that allows us to give weight to the relative riskiness of options without abandoning a broadly Bayesian framework. For instance, Stefansson and Bradley (2019) present a way to accommodate different risk attitudes in the subjective Bayesian framework due to Jeffrey (1965/1983), whereby utility is defined not only over ordinary outcome propositions, but also chance propositions (describing, e.g., a 10% chance of losing 18 QALYs). For agents who are not risk neutral, the relationship between the utilities of outcome propositions and the corresponding chance propositions will not be linear in probabilities, and risk attitudes will be captured by the nature of the non-linear relationship between these utilities – regarding which Stefansson and Bradley are permissive. Weirich (2020), on the other hand, proposes that utility can be ascribed to an act’s riskiness in a way that is separable from, and can be added to the expected utility of its outcomes as traditionally construed.² While Weirich rules out risk-seeking attitudes, he treats different levels of risk aversion as rationally permissible.

Whichever way we choose to formally accommodate permissiveness of rationality under risk, acceptance of such permissiveness opens up a number of important and difficult philosophical and practical questions. The purpose of this paper is to raise one such question, namely, what risk attitudes we should implement when choosing on behalf of another person. I will elaborate on this question and explain its wider significance (Section 2), present some existing answers to the question in this young literature (Section 3), and suggest some ways in which it can be more systematically addressed (Section 4).

²Goldschmidt and Nissan-Rozen (2020) show how such a separation is also possible in Stefansson and Bradley’s more permissive framework.

2 The Question: Risky Choice on Behalf of Others

Those who think that rationality is permissive under risk hold that, in the case introduced in the last section, you are free to prefer and choose either *Treatment 1* or *Treatment 2*, with your choice expressing your individual attitude to risk. But what if you are making that choice on behalf of another person, for instance as a next-of-kin, a medical professional, or by making funding decisions as a health policy-maker? How should you choose then? More generally, the question I wish to discuss is how we should make risky choices on behalf of other people in the context of permissiveness of rationality under risk. As a core case, I will focus on choices on behalf of a single other person (henceforth the ‘patient’) which affect only them, where both the person’s evaluation of outcomes and her risk attitude are known. I will grant that the decision-maker evaluates outcomes in the same way as the patient, either because she takes herself to be required to defer to the patient’s evaluation of outcomes, or because she takes them to be correct. I will also assume that patient and decision-maker agree on the probabilities of the various outcomes the different available options might bring about.

The problem raised by accepting permissiveness of rationality under risk is that given such permissiveness, outcome-valuations and probabilities do not on their own settle how a decision-maker should choose on behalf of a patient. It is not enough, as is often assumed, to determine which outcomes to pursue on the patient’s behalf, what their relative importance is and how likely they are to come about given different choices. Settling on outcome-valuations and probabilities is already both philosophically controversial and in many cases practically difficult. But if rationality is permissive under risk, the decision-maker also needs to settle on what attitudes to risk she should implement to arrive at a decision. I take there to be three main kinds of plausible answers to this question:

1. When choosing on behalf of another person, you are permitted to implement any rationally permissible risk attitude. You are, for instance, free to choose in accordance with your own attitude to risk (call this *permissive*).
2. When choosing on behalf of another person, there is a specific attitude to risk you are required to implement, irrespective of what your own, or the patient’s attitude to risk is. For instance, you are required to be risk neutral, or to implement the most risk averse of the reasonable and rationally permissible risk attitudes (call this *required*).
3. When choosing on behalf of another person, you should defer to the patient’s own risk attitude where known, or your best guess thereof after having taken reasonable steps to find out their risk attitude (call this *deferential*).

There can, of course, be intermediate positions between these, and which of these answers is most plausible may also depend on more specific features of the case at hand and the types of valuable outcomes at stake. For instance, when speaking of making decisions

“on behalf of” another, we may either be referring to circumstances where we make a choice that would ordinarily be made by the person on whose behalf we are choosing, as a kind of proxy decision-maker who may even have been given specific instructions, e.g. when we make medical decisions on behalf of a family member who is unconscious; or we may have in mind circumstances where we make choices that affect others, aiming to serve their interests, but that for some reason were never in the direct sphere of agency of the patient(s), e.g. when a policy-maker decides on rules and regulations, or on which projects to fund. Requirements of deference may well be stronger in the former case (more on this below).

The answer to the question of how we should make risky choices on behalf of others has important practical and theoretical implications in a variety of different contexts. Before looking at how the question has been addressed, and how I think it can be more systematically addressed, I’d like to point at just some of these. For one, the choice between *permissive*, *required* and *deferential* is highly relevant in social choice theory, which explores how a social decision-maker should choose on behalf of groups of people. A number of recent contributions to this field explore whether and how permissiveness of rationality under risk can be accommodated. If we think *deferential* is the correct approach in single-patient contexts, it seems highly plausible that this implies a commitment to a version of ex ante Pareto condition in the many-patient contexts explored by social choice theory that is deferential to each person’s risk attitude – that is, roughly, the condition that if a risky social gamble is preferred by each individual, it should also be preferred by the social decision-maker. But it has been shown that different versions of such risk attitude-deferential ex ante Pareto conditions are incompatible with various other desirable features of social choice, such as that the social decision-maker should prefer a gamble that is sure to bring about a better outcome (see Blesseohl 2020, Nebel 2020, Nissan-Rozen 2020, Bradley 2021). As Blesseohl (2020) highlights, these conflicts arise even in cases where standard arguments against ex ante Pareto (such as Fleurbaey and Voorhoeve’s 2013 argument from the value of ex post equality) do not apply. But if deference to individual risk attitudes was not desirable or required in the first place, and if the same (i.e. a risk neutral) risk attitude could be applied for all individuals, some of these problematic implications can potentially be avoided.

Secondly, how we should make risky choices on behalf of others is also relevant for a variety of debates in moral philosophy. For instance, in the contractualist tradition, broadly speaking, what matters is that the rules we live by are justifiable to each (see Ashford and Mulgan (2018) for a helpful overview). There is a debate amongst contractualists about whether in contexts of risk, it is the risky gambles that the chosen rules impose on society that should be *ex ante* justifiable to each, or it is the *ex post* outcomes of those rules that need to be justifiable to each. For ex ante contractualists, such as James (2012) or Frick (2015), once we allow for permissiveness of rationality under risk, a crucial question becomes whether ex ante justifiability requires us, in the spirit of *deferential*, to honour each individual’s attitude to risk. Scharding (2021) recently argued that this would be unworkable, but proposes that public risks are justifiable just in case they are deemed acceptable by the average risk attitude in the population.

Finally, a practical context in which the answer to the question of how we should make risky choices on behalf of others is crucial is current practice in cost-benefit analysis as a method of policy evaluation under risk. Cost-benefit analysis traditionally aims to respect the preferences of the individuals affected by a policy in its accounting of costs and benefits, and it standardly implements expected utility theory (see Boadway 2016 for a helpful introduction). But depending on how expected utility theory is implemented, the resulting cost-benefit analysis may or may not be broadly in line with *deferential* in the way in which it takes each person’s interests into account. Cost-benefit analysis can either measure the preferences of affected individuals over the risky gambles a policy will confront them with and aim to respect those, in the spirit of *deferential*. Or it may develop a cardinal measure of the degree to which the affected individuals value the potential outcomes of a policy, use that as a utility function, and take the expectation of that measure to represent the individuals’ interests in risky contexts. In health economics, for instance, QALYs are often measured in riskless contexts and used in this way (see Abellan-Perpinan et al. 2016). Those who advocate for the use of subjective wellbeing measures in cost-benefit analysis also take this approach (see Clark 2016). What this amounts to is implementing risk neutrality in choice on behalf of others irrespective of what individuals’ actual risk attitudes are, and so would only be condoned by *permissive* or a risk neutral version of *required*. Similarly, as discussed in Thoma (2021), popular methods of preference debiasing in behavioural welfare economics impose risk neutrality on individuals. What is crucial in evaluating these methods is whether the lack of deference to non-neutral risk attitudes is problematic or not.

3 Existing Answers in the Literature

A number of authors have explicitly taken a stance on the question of how to make risky choices on behalf of others once we grant permissiveness about attitudes to risk. In our core case, a decision-maker chooses on behalf of a single patient whose risk attitudes are known. In such cases, Buchak (2017a) takes it to be intuitive that “we take ourselves to be required to defer to” the patient’s risk attitude (p.632). Bovens (2015a), on the other hand, claims that “[i]t is perfectly reasonable for a person to choose more conservatively for other people than these people would choose for themselves even assuming that the choices of these people would be ideally rational and fully informed.” (p.404) It thus seems clear that direct appeal to intuition will not get us far.

Looking at the empirical literature on risky choice on behalf of others might be instructive, and here, a meta-analysis found that “people’s predictions of others’ decisions are relatively unrelated to decisions they make for others.” (Polman and Wu 2020, p.15) This suggests that in practice, people don’t tend to adhere to a strong norm of deference to others’ risk attitudes. Neither, however, do they systematically choose in a more risk averse manner, or consistently choose in line with the risk attitude they apply in self-regarding choices. In some contexts, a “cautious shift” can be observed, and in others a “risky shift”. It thus appears that it is only *permissive* that does not imply that people

regularly don't live up to their obligations when making risky choices on behalf of others.

Some potential justifications for *permissive*, *required* and *deferential* have also been offered that go beyond appeal to intuition. Starting with *deferential*, we find two main lines of argument. The first is based on the claim that a person's risk attitudes co-determine what gambles it is in a person's interest to be subjected to (see, for instance, Buchak 2017b). In our example, the claim would be that receiving *Treatment 1* would be in the interest of a risk averse patient, but not in the interest of a risk neutral patient. On this view, people's interests in risky contexts may differ even if they value outcomes in the same way. If we think that, more generally, when choosing on behalf of somebody else, we should do what is in their interests, that provides us with a justification for *deference*. The second common line of argument in favour of *deferential* is that respect for autonomy requires us to defer to what a patient would choose for themselves, with the potential caveat that their choices are not clearly opposed to their own interests (see, e.g., Jonker 2020). Hecht's (2021) claim that non-deference would be objectionably paternalistic can also be understood along those lines. This is also a familiar non-consequentialist justification for ex ante Pareto, as pointed out by Nissan-Rozen (2020).

In favour of *permissive*, Bovens (2015a) appeals to the sense of responsibility decision-makers typically feel for choices they make on behalf of others, and the expected emotional effects on decision-makers when things turn out badly for the patient. According to Bovens, this may make it reasonable for the decision-maker to make more risk averse choices than the patient herself would make. The decision-maker may legitimately say, "I cannot afford running the risk of having such bad outcomes happen on my watch." (p.404) Bovens's remarks fit well with the potential psychological explanations of "cautious shifts" when choosing for others canvassed by Polman and Wu (2020), which appeal to the greater strength of negative emotions when causing bad outcomes for others than when causing good ones. Note, however, that the more natural way to take such negative emotional effects into account would be to include them in the evaluation of outcomes. But that can't be what Bovens has in mind, since firstly, this would mean that the choice is no longer purely on behalf of another, and secondly, Bovens aims to make a point about permissible risk attitudes rather than outcome valuations. Bovens's point can't be only about *feeling* responsible and the emotional cost of that feeling. The point must be that such feelings, and their intuitive fittingness, are an indication that what risk attitude to apply genuinely falls under the responsibility of the decision-maker, that this is up to her. Building on this point, however, if we do think that risk attitudes are a decision-maker's prerogative, it is also no longer clear why we should only take attitudes that are more risk averse than the patient's to be reasonable, and not also ones that are more risk-seeking.³

Required appears to be explicitly defended mainly for contexts where the patient's risk attitudes are unknown (and presumably, cannot be easily ascertained) so that *deferential*

³Drawing on Hecht's (2021) argument below (intended, by Hecht, only to apply when the patient's risk attitude is unknown), a decision-maker may claim, in certain circumstances involving vulnerable patients, "I cannot afford for it to happen on my watch that you are deprived of having at least the chance of a minimally decent life."

is not available. Buchak's (2017) 'risk principle' (p.632) holds that we should be deferential to the patient's risk attitudes when known, and when they are unknown, apply the most risk averse of the reasonable risk attitudes (where the set of reasonable risk attitudes is understood to be smaller than the set of rationally permissible ones, ruling out extreme risk aversion and risk-seeking behaviours). The risk principle is crucial to Buchak's justification of prioritarianism, which relies on risk averse choice behind a Harsanyi- and Rawls-inspired veil of ignorance,⁴ as well as to her defence of a very risk averse response to the climate crisis (see Buchak 2018). In defence of the requirement to be risk averse when the patient's risk attitudes are unknown, Buchak claims that no reasonable person could fault a decision-maker for making, within the bounds of the reasonable, too risk averse a choice on their behalf. At the same time, a decision-maker could be faulted for being more risk-seeking than the most risk averse reasonable person. However, Buchak (2017a) does not elaborate on what justifies this asymmetry. Another potential justification for the 'risk principle', offered by Buchak (2021) more recently, is that choosing on behalf of another in the context of risk is a problem of distributional justice amongst the patient's possible future selves, whereby a concern for the worst-off possible future selves leads to risk averse choice. Indeed this is also how Bovens (2015a) characterises risk aversion when choosing on behalf of others (also see his Bovens 2015b and Bovens 2019). However, in this case, more justification is needed why this distributional argument should not extend to justify risk aversion in contexts where the patient's risk attitudes are known.⁵ A further challenge to the risk principle is offered by Hecht (2021), who builds on Hayenhjelm's (2006) discussion of taking risks from a position of vulnerability to argue that risk-seeking behaviour on behalf of others can be permissible, just in case there is a chance a risky gamble may take a patient above the threshold of entitlement to a minimally decent life. Jonker (2020), on the other hand, claims that the vulnerability of patients (e.g. in development policy) provides us with especially strong reasons to implement risk aversion where risk attitudes are not known.

For cases where there is more than one patient, results in social choice theory may also be appealed to in order to argue for or against various views on what risk attitudes should be implemented when choosing on behalf of others. For instance, short of arguing for a version of *required*, Nissan-Rozen and Fiat (2021) present a result in social choice theory to offer some support for the claim that choice on behalf of others should at least sometimes be restricted to implementing a single risk attitude that is either risk neutral or risk averse, since this is the only way of avoiding violation of two plausible principles of social choice. However, they acknowledge that, where risk attitudes are known, there may be requirements of deference that could undermine this argument.

⁴Also see Stefansson (2021) who treats ambiguity aversion as another kind of risk aversion and extends the 'risk principle' to include it, to derive egalitarianism from behind the veil of ignorance.

⁵Moreover, this justification for the risk principle seems to be question-begging for the purposes of Buchak's (2017a) derivation of a prioritarianism from behind the veil of ignorance.

4 Some Thoughts on a Way Forward

While the choice between *permissive*, *required* and *deferential* is of much importance for a number of practical and theoretical purposes, the last section has shown there is neither convergence on any one view, nor are any of the existing justifications for the different views conclusive. What appears to be particularly controversial is whether any requirements of deference to a patient's own values or choices extends to risk attitudes. If yes, this speaks in favour of *deferential*. If no, there is a prima facie case for extending the permissiveness of rationality in self-regarding choices to choices on behalf of others. But the absence of a requirement of deference would also potentially open the door for moral arguments to restrict permissible risk attitudes in other-regarding contexts, e.g. based on impossibility results from social choice theory in many-patient cases.

To develop a more principled stance on how we should take risks on behalf of others, the question of whether deference to another's risk attitudes is required is thus crucial. To address this question, I believe it is helpful to consider two more fundamental issues: First, what kind of deference to patients' values, judgements and/or choices do we take to be desirable in riskless contexts, and why? Those who believe there are no requirements of deference to patients' values and judgements even in riskless contexts will clearly not find *deference* attractive for risk attitudes either, and instead accept either *permissive* or *required*. But those who accept there are reasons for deference in riskless contexts may inquire into whether these reasons extend to risk attitudes. For that, a second question is crucial: What kinds of attitudes are risk attitudes? What is their moral significance?

To start with the first question, we can distinguish between broadly consequentialist and broadly non-consequentialist justifications for deference to a patient's own values, judgements and/or choices in riskless contexts. On the consequentialist side, the core idea is that deferring to an agent is the best or the only way to bring about what is good for her. This could be either because one thinks that the agent's own attitudes, her desires and/or judgements, constitute what is good for her (see the discussion of desire theories of wellbeing in Crisp 2017). Or it could be because one thinks that an agent's own desires, judgements or choices are the best evidence we have of what is good for her.⁶ On the non-consequentialist side, the general idea is that there is a duty or obligation a decision-maker has to defer to the patient. A broadly Kantian thought common in discussions of what is wrong with paternalism is that deference is required in order to respect the patient's autonomy.⁷ This idea is most plausible in cases where the decision-maker acts as a proxy in contexts where ordinarily the patient would make her own choices. In some contexts of proxy choice, decision-makers are even given concrete instructions of how to choose on behalf of a patient. By taking on the role of the proxy, decision-makers appear to take on an obligation to abide by these instructions. In contexts where a policy-maker decides

⁶See Hausman and McPherson (2009) on the view that this explains deference to preferences in welfare economics, and Mill's (1859) classic argument that policy-makers are likely to get things wrong when they don't defer to people on what is good for them.

⁷See, e.g. Groll (2012), and, in the context of welfare economics, Sugden (2018).

on behalf of the community on matters only the policy-maker can decide on, decision-makers are not proxy agents in this way. But here another non-consequentialist reason for deference is often cited, namely a norm of liberal neutrality, requiring that the state remain neutral on different conceptions of the good life. Deferring to citizens' own values is a way of maintaining such neutrality.⁸

With these different views on the point of deference on the table, the question for us is whether these extend to deference to risk attitudes, or whether they apply only to patients' attitudes to outcomes. To answer that question, we need to get clearer on what kinds of attitudes risk attitudes are. Amongst those who believe rationality is permissive under risk, there is disagreement about how to think of the nature of the different permissible risk attitudes one might have. An important dividing line is between those who think that attitudes to risk are merely instrumental, and those who think they also capture how agents intrinsically value features of gambles. Buchak (2013) falls in the first camp. In her risk-weighted expected utility theory, utilities assigned to outcomes are meant to fully capture an agent's ultimate ends, that is, all of the things she values non-instrumentally. Risk attitudes, on her interpretation, capture how an agent structures the attainment of her ends. Permissiveness of rationality under risk, for her, is permissiveness of instrumental rationality: There are different permissible ways to structure the attainment of one's ends under risk, and different risk attitudes capture an agent's approach to choosing amongst different permissible means.

Stefansson and Bradley (2019), Weirich (2020) and Goldschmidt and Nissan-Rozen (2020), on the other hand, explicitly reject this picture. On Stefansson and Bradley's (2019) interpretation of their framework, attitudes to risk capture desires that differ from desires for outcomes not in kind, but merely in object, in being about chances. Risk attitudes are thus not merely instrumental, but rather capture how agents intrinsically, non-instrumentally value features of gambles. Permissiveness of rationality under risk, on their interpretation, is not about allowing different means to pursuing one's ends. Rather, it is permissiveness about how the way one values chances may relate to how one values outcomes. Weirich (2020) takes the instrumental value of risky gambles to be expectational, and accommodates risk aversion only as a way of intrinsically dis-valuing the riskiness of gambles.

To illustrate the difference between the merely instrumental and the non-instrumental way of thinking about risk attitudes, suppose a decision-maker defers to your outcome-valuations, but chooses on your behalf in a less risk averse manner than you would have done yourself, for instance by choosing *Treatment 2* above when you would have picked *Treatment 1*. You are lucky, and the best potential outcome comes about. Is there any sense in which your ends were not served? If risk attitudes are merely instrumental, the answer is 'no' – you would have gone about things differently, but you got what you wanted in every sense that matters. If risk attitudes are non-instrumental, the answer is 'yes' –

⁸See Dimock (2000) for helpful general discussion of the concept of liberal neutrality and Heath (2020) on the idea that the deference implemented in standard cost-benefit analysis is best understood as an expression of liberal neutrality.

nevermind the outcome, you were deprived of chances you intrinsically value. Notably, the major existing permissive frameworks adopt only one or the other interpretation of attitudes to risk, and correspondingly only allow for permissiveness regarding one or the other type of attitude. I will thus focus on these exclusive views. But note that it is possible both play a role in the psychological explanation of people’s actual attitudes to risk,⁹ and it is not obvious why rationality should then not be permissive about both.

If one accommodates and is permissive about only non-instrumental risk attitudes, then it becomes very plausible that any argument for deference to attitudes regarding outcomes would extend to risk attitudes. If the things I non-instrumentally value constitute what is good for me, or if my valuations are the best evidence of what is good for me, then risk attitudes constitute or provide evidence of what risky gambles are good for me. This fits with remarks by some of the authors discussed in the last section that risk attitudes are partly constitutive of an agents’ interests. Moreover, there would be no reason to think that requirements of respect for patient autonomy or liberal neutrality should apply to valuations of outcomes but not risk attitudes.

Things are less straightforward, however, when we accommodate and are permissive about only instrumental attitudes to risk. Suppose a decision-maker defers to a patient’s attitudes to outcomes, but not her risk attitudes. For instance, take again a decision-maker who picks *Treatment 2* for a risk averse patient. If risk attitudes are merely instrumental, then this is a permissible way to pursue the patient’s own ends (QALYs in this case). It merely structures the pursuit of those ends differently from how the patient herself would have pursued them. On consequentialist views, it is hard to see what further deference could be required. It is an agent’s ends, her non-instrumental valuations, that are typically taken by those consequentialists to constitute or provide evidence of what is good for her. As long as the decision-maker defers to the patient’s attitudes to outcomes, and applies a risk attitude within the bounds of the rational and reasonable, she is then genuinely pursuing the patient’s good.¹⁰

This leaves open whether according to non-consequentialist views, respect for autonomy or liberal neutrality extend to risk attitudes, even if these are merely instrumental. Perhaps we should respect the autonomous choices people make not only about what ends to pursue, but also about how they wish to pursue them. And perhaps a policy-maker should not impose any particular way of pursuing their ends on people. Whether this is plausible, I think, depends on whether we think risk attitudes are, in Nissan-Rozen’s (2020) words, a “morally significant feature of a person” (p.13) or not. Suppose you have to pick between two heads of lettuce at the supermarket that appear identical to you in all important respects. They would each serve your lettuce-eating ends equally well. You pick the left one. In fact, you have developed, arbitrarily, a habit of always picking the left of two equivalent options that are spatially arranged left-to-right. We can think of your

⁹See Cohen et al. (2021) for some potential evidence that this is so.

¹⁰In line with this, Nebel and Buchak (2021) have recently developed an account of goodness of gambles for individuals according to which, when all reasonable risk attitudes don’t agree on a ranking of two gambles, it is indeterminate which is better for the patient, irrespective of her own risk attitude.

disposition to choose the left of two equivalent options as a merely instrumental attitude, one that picks out one out of the permissible means of pursuing your ends. But this attitude does not seem morally significant. It does not seem like the kind of attitude it would be disrespectful or illiberal for a (public) decision-maker to override, unless acting in a proxy capacity with clear instructions to implement the left-picking strategy (and even then, one might think this is an unreasonable demand one's obligations of deference don't extend to). It's not even the kind of thing the decision-maker herself has any good reason to abide by (or to instruct a proxy decision-maker to abide by). If risk attitudes were just such largely arbitrary habits of picking amongst permissible means towards one's ends, there also would not be an intuitive non-consequentialist case for a general requirement of deference to risk attitudes.

Could risk attitudes be merely instrumental, but morally more significant than that? Perhaps risk attitudes are character traits that are important to an agent's identity, and this grounds a requirement of deference. Or perhaps, as argued by Murray and Buchak (2019), they are volitional states, like plans or resolutions, representing how an agent has made up her mind to weight chances of better and worse outcomes. It is often thought that there is some rational pressure for an agent herself to act in accordance with such volitional states. Perhaps this also makes it the case that a decision-maker acting on behalf of a patient should honour the patient's risk attitudes, thought of as merely instrumental volitional states. Aside from the difficulty of explaining how risk attitudes can be merely instrumental yet morally significant, there is a further worry about this kind of attempt to ground a requirement of deference to risk attitudes: While it is clear that any patient would generally want a decision-maker acting on their behalf to defer to and pursue their ends, it is not obvious that patients would generally want decision-makers to adopt their merely instrumental risk attitudes when acting on their behalf.¹¹ Perhaps I have a cautious character, but I would like or tolerate somebody else who acts on my behalf to be more daring. Or perhaps I am resolved to structure the attainment of my goals in risky contexts in a certain way, but take that resolution to strictly only pertain to cases where I am the chooser, and to be irrelevant to how my interests are pursued when I am not the chooser.¹² This seems to both undermine non-consequentialist arguments for deferring to risk attitudes in the absence of clear instructions to do so, and also suggests that even where proxy agents are given clear instructions, these may not be to implement the patient's risk attitude.

What we have found, then, is that standard arguments in favour of deference plausibly extend to risk attitudes when we take risk attitudes to express only non-instrumental ways of valuing features of gambles. If we take risk attitudes to be merely instrumental, however, the case for deference is not clear. In fact, consequentialist arguments for deference appear to be inapplicable to risk attitudes under that interpretation, and the prospects of non-consequentialist arguments for deference to merely instrumental risk attitudes are

¹¹I thank Ittay Nissan-Rozen for raising this point in conversation.

¹²In Bratman's recent work on planning agency, for instance, rational pressure to act in accordance with one's volitional states is grounded in the goal of self-governance, which is a goal that does not seem to apply when others act on your behalf. See Bratman (2017).

also doubtful. To determine how we should make risky choices on behalf of others, it is thus crucial to settle the question of what kinds of attitudes the risk attitudes are that, according to the recent decision-theoretic literature, rationality is permissive about. Indeed, settling that question turns out to be more important than settling on a specific account of what is desirable about deference more generally, if it is at all desirable. If we were, moreover, to move to a framework that accommodates both instrumental and non-instrumental types of risk aversion and is permissive about both, the question of how to make risky choices on behalf of others would be further complicated, given requirements of deference may apply to one but not the other, but they will be difficult to tell apart in practice.

5 Conclusion and Outlook

The question of how to make risky choices on behalf of others is complicated by accepting, as many decision theorists have recently done, that rationality is permissive under risk, allowing for more or less risk averse, or even risk-seeking pursuit of good outcomes. Does this permissiveness of rationality imply a permission for decision-makers to apply any reasonable risk attitude when choosing on behalf of a patient? Is there, instead, some specific risk attitude decision-makers are required to implement whenever choosing on behalf of another? Or are decision-makers required to defer to the patient's risk attitudes where known? I have shown that there is little agreement in the literature so far on the right answer to these questions, and argued that to answer them, it is important to settle on an account of what kinds of attitudes the risk attitudes are rationality is permissive about.

For those who accept the permissiveness of rationality under risk, answering the questions I have outlined here is clearly of great theoretical and practical importance. Take just these two examples mentioned earlier: first, we saw that, in social choice theory, *deferential* is often appealed to when defending the intuitive pull of an ex ante Pareto condition in terms of people's preferences over risky social gambles. If *deferential* is not required, as we saw is plausible if we take risk attitudes to be merely instrumental, such a deferential ex ante Pareto condition could be relaxed without cost. Given such a condition is incompatible with various other desirable conditions of social choice, this would be good news. Giving up ex ante Pareto would be more difficult when we take risk attitudes to be non-instrumental, as then the standard arguments in favour of deference to an agent's values and choices, if we take them to have any weight at all, seem to extend to risk attitudes. And second, methods of policy evaluation that impose risk neutrality no matter what people's actual risk attitudes are are problematic if *deferential* is required. But they may be vindicated if *deferential* is not required – though they may still face criticism from those advocating for a requirement for public decision-makers to be risk averse rather than risk neutral. This article has aimed to make some contribution to progress in these debates.

References

- Jose-Maria Abellan-Perpinan, Carmen Herrero, and Jose-Luis Pinto-Prades. QALY-based cost-effectiveness analysis. In Matthew D. Adler and Marc Fleurbaey, editors, *The Oxford Handbook of Well-Being and Public Policy*. Oxford University Press, 2016.
- Elizabeth Ashford and Tim Mulgan. Contractualism. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, summer 2018 edition, 2018.
- Simon Bleszenohl. Risk attitudes and social choice. *Ethics*, 130(4):485–513, 2020.
- Robin Boadway. Cost-benefit analysis. In Matthew D. Adler and Marc Fleurbaey, editors, *The Oxford Handbook of Well-Being and Public Policy*. Oxford University Press, 2016.
- Luc Bovens. Concerns for the poorly off in ordering risky prospects. *Economics and Philosophy*, 31(3):397–429, 2015a.
- Luc Bovens. Evaluating risky prospects: the distribution view. *Analysis*, 75(2):243–253, 2015b.
- Luc Bovens. The ethics of making risky decisions for others. In Mark D. White, editor, *The Oxford Handbook of Ethics and Economics*. Oxford University Press, 2019.
- Richard Bradley. Impartial evaluation under ambiguity. Unpublished manuscript, 2021.
- Michael Bratman. Rational planning agency. In Anthony O’Hear, editor, *Philosophy of Action, Royal Institute of Philosophy Supplement*, volume 80, pages 25–48. Cambridge University Press, 2017.
- John Broome. *Weighing Goods*. Blackwell, 1991.
- Lara Buchak. *Risk and Rationality*. Oxford University Press, 2013.
- Lara Buchak. Taking risks behind the veil of ignorance. *Ethics*, 127(3):610–644, 2017a.
- Lara Buchak. 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, 2017b.
- Lara Buchak. Weighing the risks of climate change. *The Monist*, 102(1):66–83, 2018.
- Lara Buchak. Risk and ambiguity in ethical decision-making. Unpublished manuscript, August 2021.
- Andrew Clark. SWB as a measure of individual well-being. In Matthew D. Adler and Marc Fleurbaey, editors, *The Oxford Handbook of Well-Being and Public Policy*. Oxford University Press, 2016.
- Haim Cohen, Anat Maril, Sun Blechier, and Ittay Nissan-Rozen. Attitudes toward risk are complicated. Unpublished manuscript, August 2021.

- Roger Crisp. Well-being. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, fall 2017 edition, 2017.
- Susan Dimock. Liberal neutrality. In Jan Narveson and Susan Dimock, editors, *Liberalism*, pages 41–58. Springer, 2000.
- Marc Fleurbaey and Alex Voorhoeve. Decide as you would with full information: An argument against ex ante Pareto. In Nir Eyal, Samia Hurst, Ole Norheim, and Dan Wikler, editors, *Inequalities in Health: Concepts, Measures, and Ethics*. Oxford University Press, 2013.
- Johann Frick. Contractualism and social risk. *Philosophy and Public Affairs*, 43:175–223, 2015.
- Zeev Goldschmidt and Ittay Nissan-Rozen. The intrinsic value of risky prospects. *Synthese*, 198(8), 2020.
- Daniel Groll. Paternalism, respect, and the will. *Ethics*, 122:692–720, 2012.
- Daniel Hausman and Michael McPherson. Preference satisfaction and welfare economics. *Economics and Philosophy*, 25:1–25, 2009.
- Madeleine Hayenhjelm. Out of the ashes: Hope and vulnerability as explanatory factors in individual risk taking. *Journal of Risk Research*, 9(3):189–204, 2006.
- Joseph Heath. *The Machinery of Government*, chapter Cost-Benefit Analysis as an Expression of Liberal Neutrality. Oxford University Press, 2020.
- Lisa Hecht. Permissible risk-inclination for benevolent decision-makers. Unpublished manuscript, May 2021.
- Aaron James. Contractualism’s (not so) slippery slope. *Legal Theory*, 18:263–292, 2012.
- Richard Jeffrey. *The Logic of Decision*. University of Chicago Press, 2nd edition, 1965/1983.
- Julian Jonker. Risk and asymmetry in development ethics. *African Journal of Business Ethics*, 14(1):23–41, 2020.
- John Stuart Mill. *On Liberty*. Cambridge University Press, 2001, 1859.
- Dylan Murray and Lara Buchak. Risk and motivation: When the will is required to determine what to do. *Philosopher’s Imprint*, 19(16):1–12, 2019.
- Jacob Nebel. Rank-weighted utilitarianism and the veil of ignorance. *Ethics*, 131:87–106, 2020.
- Jacob Nebel and Lara Buchak. The value of risky prospects. Unpublished handout, May 2021.

- Ittay Nissan-Rozen. Weighing and aggregating reasons under uncertainty: a trilemma. *Philosophical Studies*, Online first: <https://doi.org/10.1007/s11098-020-01587-9>, 2020.
- Ittay Nissan-Rozen and Jonathan Fiat. Attitudes to risk when choosing for others. Unpublished manuscript, 2021.
- Evan Polman and Kaiyang Wu. Decision making for others involving risk: A review and meta-analysis. *Journal of Economic Psychology*, 77(C), 2020.
- John Quiggin. A theory of anticipated utility. *Journal of Economic Behavior & Organization*, 3(4):323–343, 1982.
- Leonard Savage. *The Foundations of Statistics*. Dover Publications, second revised edition, 1972.
- Tobey K. Scharding. Contractualism and risk preferences. *Economics and Philosophy*, 37(2):260–283, 2021.
- H. Orri Stefansson. Ambiguity aversion behind the veil of ignorance. *Synthese*, 198:6159–6182, 2021.
- H. Orri Stefansson and Richard Bradley. What is risk aversion? *British Journal for the Philosophy of Science*, 70(1):77–102, 2019.
- Robert Sugden. *The Community of Advantage: A Behavioural Economist’s Defence of the Market*. Oxford University Press, 2018.
- Johanna Thoma. Merely means paternalist? Prospect theory and ‘debiased’ welfare analysis. Unpublished manuscript, April 2021.
- David Velleman. The story of rational action. In *The Possibility of Practical Reason*. Oxford University Press, 1993/2000.
- John von Neumann and Oskar Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, 1944.
- Milton C. Weinstein, George Torrance, and Alistair McGuire. QALYs: The basics. *Value in Health*, 12(s1):S5–S9, 2009.
- Paul Weirich. *Rational Responses to Risk*. Oxford University Press, 2020.