

ARTICLE

Taking Risks on Behalf of Another

Johanna Thoma 

Department of Philosophy, Logic and Scientific Method, London School of Economics and Political Science, London, UK

Correspondence

Johanna Thoma, Department of Philosophy, Logic and Scientific Method, London School of Economics and Political Science, London, UK.
Email: j.m.thoma@lse.ac.uk

Abstract

A growing number of decision theorists have, in recent years, defended the view that rationality is permissive under risk: Different rational agents may be more or less risk-averse or risk-inclined. This can result in them making different choices under risk even if they value outcomes in exactly the same way. One pressing question that arises once we grant such permissiveness is what attitude to risk we should implement when choosing on behalf of other people. 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 people 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. Consider a decision problem under risk: There is an agent who faces a choice between a number of different options. She values the different potential outcomes of her choice to different extents, but she does not know for sure what outcome each of the options available to her will bring about. She can, however, assign probabilities to how likely different outcomes are to come about as a consequence of the available options. Those who believe that rationality is permissive under risk grant that rational agents are free to make such risky choices in more or less risk-averse or risk-inclined ways. If rationality is permissive under risk, then the probabilities assigned to

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2023 The Authors. Philosophy Compass published by John Wiley & Sons Ltd.

the potential outcomes, along with knowing the extent to which the agent values the outcomes do not fix how she should rationally choose.

The following example illustrates this idea. Suppose you have to choose between two different medical treatments to undergo for a life-threatening condition. These treatments do not affect the quality of your life, and you moreover judge each of your remaining years of life to be equally valuable. However, the treatments affect how many years of life you have left to live. On the best available evidence, the two treatments available to you have the following prospects:

Treatment 1: Lose 2 life-years for certain.

Treatment 2: 10% chance of losing 18 life-years (and none otherwise).

Which treatment should you choose? A risk-neutral agent would maximise the expectation of life-years, 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 life-years. *Treatment 1* is a more risk-averse course of action. Intuitively, both the risk-neutral and the risk-averse choice seem rationally permissible. To accommodate this intuition is to accept that rationality is permissive under risk.

In real-world cases, of course, the quality of our life changes over time, and medical treatments affect the quality of life and not just the number of life-years left. Consequently, the metric commonly used for evaluating the cost effectiveness of health interventions is quality-adjusted life-years (QALYs). QALYs are intended to be a cardinal measure of people's health-related values: Losing two QALYs is twice as bad as losing one, and so on. 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. If QALY measures are successful in their ambition, then they provide a useful illustration of the permissiveness thesis – rationality may be permissive about whether agents are risk-averse, risk-neutral, or potentially even risk-inclined in their pursuit of QALYs. We can consider, for instance, a version of the example just given involving QALYs rather than life-years. And their wide-spread use in practice also provides a test case of the practical implications of accepting the permissiveness thesis to be explored in what follows.¹

Permissiveness under risk can be accommodated in a variety of different ways in formal decision theories. Let me briefly explain three approaches here. The first approach is to accommodate it within expected utility theory (EUT) in the von Neumann and Morgenstern (1944) and Savage (1972) traditions. Next to probabilities, EUT in these traditions 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 utilities.² On one natural way of understanding this, the theory requires risk neutrality, and thus cannot accommodate permissiveness. If utility is just understood to be a measure of the extent to which agents value outcomes,³ EUT requires agents to pursue value in a risk-neutral way. EUT is often understood in this risk-neutral way in the philosophical literature, including by moral philosophers concerned with the question of how to make risky choices on behalf of others (e.g. Otsuka, 2015; Otsuka & Voorhoeve, 2009).⁴

However, EUT can potentially accommodate permissiveness if we give up on the idea that utility must be understood as a cardinal measure of an agent's valuation of outcomes. In our example, we want to accommodate risk aversion regarding life-years, despite the agent valuing each life-year to the same extent. A true cardinal measure of the extent to which the agent values life-years would be linear in life-years; but we can only accommodate risk aversion in EUT if utility is not linear in life-years. Instead, to accommodate risk aversion, we must assign the agent decreasing marginal utility in life-years. More generally, to accommodate risk aversion or risk-seeking regarding value itself within EUT in the von Neumann-Morgenstern and Savage traditions, the shape of the utility function must in part capture the agent's risk attitudes, on top of capturing her valuation of outcomes. Permissiveness of rationality under risk finds expression in permissiveness about the shape of the utility function holding fixed valuation of outcomes. In contrast to other ways of accommodating permissiveness, this strategy does not relax the key axioms of the representation theorems of EUT. This is an advantage if one finds these axioms normatively compelling; however it also means 'anomalous' risk-averse preference structures, like those in the Allais puzzle (Allais, 1953) cannot be accommodated.

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.⁵ This allows agents to give greater or smaller weight to especially bad potential outcomes, capturing greater or smaller levels of risk aversion. On this approach, permissiveness of rationality under risk amounts to permissiveness about the risk function. One potential advantage of this approach over accommodation within EUT is that when risk attitudes are captured by the risk function, utility can again function as a cardinal measure of the value of outcomes. Moreover, since they relax some of the core axioms of EUT, various 'anomalous' risk-averse preference structures, like the Allais ones, can be accommodated in non-expected utility theories (as demonstrated, e.g. in Buchak, 2013).⁶

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). In their approach, utility is defined not only over ordinary outcome propositions, but also chance propositions (describing, e.g., a 10% chance of losing 18 life-years). 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. Risk attitudes are then 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.⁷ While Weirich rules out risk-seeking attitudes, he treats different levels of risk aversion as rationally permissible.

These different ways of formally accommodating permissiveness of rationality under risk differ both in just how permissive they are, and in what interpretation of risk aversion they most plausibly capture (more on this below). But 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 other people. 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).

2 | THE QUESTION: RISKY CHOICE ON BEHALF OF OTHERS

Those who think that rationality is permissive under risk will likely hold that, in the case introduced in the last section, you are free to prefer and choose either *Treatment 1* or *Treatment 2*. Your choice expresses your individual attitude to risk. But what if you are making that choice on behalf of another person or other people, 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, and to bracket issues of aggregation, I will focus on choices on behalf of a single other person (henceforth the 'patient'). But the different views on these single-patient cases that I will discuss have implications for how each patient's interests should be taken into account in multi-patient contexts, which I will comment on throughout. I will moreover focus on cases where the patient'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 for those who accept that there is more than one rationally permissible attitude to risk⁸:

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. 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 have different things in mind. We may 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. 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.

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 the following version of an ex ante Pareto condition in the many-patient contexts explored by social choice theory: If every patient prefers (ex ante, so before uncertainty is resolved) risky social choice *a* to risky social choice *b*, then the social decision-maker should also prefer *a* to *b*. Abiding by this ex ante Pareto condition would amount to showing deference to each person's attitude to risk, even if it is non-neutral. If rationality is permissive under risk, then even an ex ante Pareto condition applied only to patients' *rational* preferences implies deference to non-neutral attitudes to risk. But it has been shown that various kinds of ex ante Pareto conditions that allow for non-neutral individual attitudes to risk 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 Blessenohl, 2020; Bradley, 2022; Nebel, 2020; Nissan-Rozen, 2020). As Blessenohl (2020) highlights, these conflicts arise even in cases where standard arguments against ex ante Pareto (such as Fleurbaey & Voorhoeve, 2013) do not apply. But if deference to individual risk attitudes was not desirable or required in the first place, and if the same (and especially a risk-neutral) risk attitude could be applied for all individuals, some of these problematic implications can 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 & 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 EUT (see Boadway, 2016 for a helpful introduction). But depending on how EUT 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 well-being measures in cost-benefit analysis also take this approach (see Clark, 2016): Subjective wellbeing measures aim to measure more directly than traditional economic choice-based methods people's own sense of how well their lives are going – either by asking about their overall life satisfaction, or eliciting moment-by-moment hedonic evaluations. Cost-benefit analysis based on these measures proceeds to maximise the expectation of subjective wellbeing. This approach amounts to implementing risk neutrality in choice on behalf of others irrespective of what individuals' actual risk attitudes are. 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 attitude is known. In such cases, views on whether deference is required differ. 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.”⁹ (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 & 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 herself, with the potential caveat that her choices are not clearly opposed to her 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 is genuinely the decision-maker's prerogative. 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.¹⁰

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* is not available. Buchak's (2017a) '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,¹¹ 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. 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.¹² 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. They argue that this is the only way of avoiding violation of at least one of two plausible principles of social choice: ex ante Pareto as I characterised it above, and the condition that a social decision-maker should not choose an option when there is another option that will certainly bring about higher total utility and a more egalitarian distribution of that utility. However, they acknowledge that, where risk attitudes are known, there may be requirements of deference that could undermine this argument.

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 consequentialist and 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 a patient is the best or the only way to bring about what is good for her. This could be either because one thinks that the patient'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 a patient's own desires, judgements or choices are the best evidence we have of what is good for her.¹³ 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.¹⁴ 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 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.¹⁵

Do these justifications of deference extend to risk attitudes, or do 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 no agreement on the nature of risk attitudes. 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 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' – 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 are correspondingly only permissive about one or the other type of attitude. But note that it is possible both play a role in the psychological explanation of people's actual attitudes to risk,¹⁶ 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 a patient's 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 attitude. 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 (life-years or 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 a patient'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.¹⁷

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 judgments people make not only about what ends to pursue, but also about how they wish them to be pursued. 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 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. It's not even the kind of thing the decision-maker herself has any good reason 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. However, it is unclear whether this rational pressure extends to a decision-maker acting on behalf of a patient.¹⁸ And there is a more general worry: While it is clear that any patient would generally want a decision-maker acting on her behalf to defer to and pursue her ends, it is not obvious that patients would generally want decision-makers to adopt their merely instrumental risk attitudes when acting on their behalf.¹⁹ Perhaps, for instance, I have a cautious character, but I would like or tolerate somebody else who acts on my behalf to be more daring.

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 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 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: requirements of deference may apply to one but not the other, but they will be difficult to tell apart in practice.

5 | CONCLUSION

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. Does this permissiveness 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 a patient's risk attitude 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 that 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 as well as practical importance.

ACKNOWLEDGMENTS

I thank participants at the LSE Protective Belt Seminar, the Online Workshop on Risk Attitudes, the University of Vienna, the University of Southampton, the Munich Centre for Mathematical Philosophy, the University of Tennessee

at Knoxville, the Pittsburgh Formal Epistemology Workshop, and the University of Stirling for very useful and interesting discussion. I am also grateful to Luc Bovens, Lara Buchak, Lisa Hecht, Ittay Nissan-Rozen, H. Orri Stefánsson, Keith Hankins and two anonymous referees for very helpful comments on previous drafts of this paper.

ORCID

Johanna Thoma  <https://orcid.org/0000-0002-1364-4521>

ENDNOTES

¹ See Weinstein et al. (2009) for an introductory overview of QALY measurement, and Bognar and Hirose (2014) for an introduction to the core ethical and methodological issues. Most methods for determining quality weights in practice are based on eliciting people's evaluations of health outcomes under conditions of certainty: A standard set of generic health states are evaluated using visual analogue scale or time trade-off methods. But note that one method popular in theory, but less common in practice derives quality-weights from people's choices or expressed preferences under risk, namely the standard gamble method. The validity of this method for providing a cardinal measure of health-related quality of life in fact depends on the assumption of risk neutrality, and is thus called into question by the permissiveness thesis.

² As a simple formal statement, act a 's expected utility is given by:

$$EU(a) = \sum_{i=1}^n p(o_i) \cdot u(o_i)$$

where there are n outcomes o_i , $u(o_i)$ is a function assigning utilities to outcomes, and $p(o_i)$ is a function assigning probabilities to outcomes.

³ This interpretation of utility is sometimes called 'realist', or 'substantive'. See Bermudez (2009), Buchak (2016), Okasha (2016), Thoma (2019a). The plausibility of risk aversion with regard to value itself is sometimes appealed to as an argument in favour of giving up this interpretation of utility in EUT, 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 (1993/2000).

⁴ There are also explicit arguments in favour of a risk-neutral version of EUT. In the moral context, see, for instance, Zhao (forthcoming).

⁵ Formally, let o_1, o_2, \dots, o_n be the outcomes ordered by their utility from lowest to highest, and let the risk-weighting function $r(p)$ be a function of probability. The REU of an act a is then given by

$$REU(a) = u(o_1) + r\left(\sum_{i=2}^n p(o_i)\right)(u(o_2) - u(o_1)) + r\left(\sum_{i=3}^n p(o_i)\right)(u(o_3) - u(o_2)) \\ + \dots + r(p(o_n))(u(o_n) - u(o_{n-1}))$$

⁶ However, also see Briggs (2015) and Pettigrew (2015) pointing out the costs of this accommodation, and Thoma (2019b) and Thoma and Weisberg (2017) calling into question whether REU theory can really accommodate anomalous preferences after all.

⁷ Goldschmidt and Nissan-Rozen (2020) show how such a separation is also possible in Stefánsson and Bradley's more permissive framework.

⁸ Note that, for those who do not take rationality to be permissive under risk, *permissive* requires agents to implement the uniquely permissible attitude to risk, and so reduces to a form of *required*. While my discussion is primarily aimed at those who accept that rationality is permissive under risk, some of the discussion also illuminates the choice between *required* and *deferential* still faced by those who reject the permissiveness of rationality under risk.

⁹ Note, however, that Buchak and Bovens concur in the need to err on the side of risk aversion when the patient's attitude to risk is not known, as discussed below.

¹⁰ Drawing 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."

¹¹ Also see Stefánsson (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.

- ¹² Moreover, this justification for the risk principle seems to be question-begging for the purposes of Buchak's (2017a) derivation of prioritarianism from behind the veil of ignorance.
- ¹³ See 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.
- ¹⁴ See, e.g. Groll (2012), and, in the context of welfare economics, Sugden (2018).
- ¹⁵ See 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.
- ¹⁶ See Cohen et al. (2022) for some potential evidence that this is so.
- ¹⁷ In 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.
- ¹⁸ In 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).
- ¹⁹ I thank Ittay Nissan-Rozen for raising this point in conversation.

REFERENCES

- Abellan-Perpinan, J.-M., Herrero, C., & Pinto-Prades, J.-L. (2016). QALY-based cost-effectiveness analysis. In M. D. Adler & M. Fleurbaey (Eds.), *The Oxford Handbook of Well-Being and Public Policy*. Oxford University Press.
- Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école américaine. *Econometrica*, 21(4), 503–546. <https://doi.org/10.2307/1907921>
- Ashford, E., & Mulgan, T. (2018). Contractualism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (summer 2018 edition). Metaphysics Research Lab, Stanford University.
- Bermudez, J. L. (2009). *Decision Theory and Rationality*. Oxford University Press.
- Blessenohl, S. (2020). Risk attitudes and social choice. *Ethics*, 130(4), 485–513. <https://doi.org/10.1086/708011>
- Bodway, R. (2016). Cost-benefit analysis. In M. D. Adler & M. Fleurbaey (Eds.), *The Oxford Handbook of Well-Being and Public Policy*. Oxford University Press.
- Bognar, G., & Hirose, I. (2014). *The Ethics of Health Care Rationing: An Introduction*. Routledge.
- Bovens, L. (2015a). Concerns for the poorly off in ordering risky prospects. *Economics and Philosophy*, 31(3), 397–429. <http://doi.org/10.1017/s0266267115000188>
- Bovens, L. (2015b). Evaluating risky prospects: the distribution view. *Analysis*, 75(2), 243–253. <https://doi.org/10.1093/analysis/avn023>
- Bovens, L. (2019). The ethics of making risky decisions for others. In M. D. White (Ed.), *The Oxford Handbook of Ethics and Economics*. Oxford University Press.
- Bradley, R. (2022). Impartial evaluation under ambiguity. *Ethics*, 132(3), 541–569. <https://doi.org/10.1086/718081>
- Bratman, M. (2017). Rational planning agency. In A. O'Hear (Ed.), *Philosophy of Action, Royal Institute of Philosophy Supplement* (Vol. 80, pp. 25–48). Cambridge University Press.
- Briggs, R. A. (2015). Costs of abandoning the sure-thing principle. *Canadian Journal of Philosophy*, 45(5–6), 827–840. <https://doi.org/10.1080/00455091.2015.1122387>
- Broome, J. (1991). *Weighing Goods*. Blackwell.
- Buchak, L. (2013). *Risk and Rationality*. Oxford University Press.
- Buchak, L. (2016). Decision theory. In A. Hajek & C. Hitchcock (Eds.), *The Oxford Handbook of Probability and Philosophy*. Oxford University Press.
- Buchak, L. (2017a). Taking risks behind the veil of ignorance. *Ethics*, 127(3), 610–644. <https://doi.org/10.1086/690070>
- Buchak, L. (2017b). 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. <https://doi.org/10.1136/medethics-2015-103118>
- Buchak, L. (2018). Weighing the risks of climate change. *The Monist*, 102(1), 66–83. <https://doi.org/10.1093/monist/ony022>
- Buchak, L. (2021). Risk and ambiguity in ethical decision-making. Unpublished manuscript.
- Clark, A. (2016). SWB as a measure of individual well-being. In M. D. Adler & M. Fleurbaey (Eds.), *The Oxford Handbook of Well-Being and Public Policy*. Oxford University Press.
- Cohen, H., Maril, A., Blechier, S., & Nissan-Rozen, I. (2022). Attitudes toward risk are complicated. *Philosophical Studies*, 179(8), 2553–2577. Online First. <https://doi.org/10.1007/s11098-022-01778-6>
- Crisp, R. (2017). Well-being. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (fall 2017 edition). Metaphysics Research Lab, Stanford University.

- Dimock, S. (2000). Liberal neutrality. In J. Narveson & S. Dimock (Eds.), *Liberalism* (pp. 41–58). Springer.
- Flourbaey, M., & Voorhoeve, A. (2013). Decide as you would with full information: An argument against ex ante Pareto. In N. Eyal, S. Hurst, O. Norheim, & D. Wikler (Eds.), *Inequalities in Health: Concepts, Measures, and Ethics*. Oxford University Press.
- Frick, J. (2015). Contractualism and social risk. *Philosophy and Public Affairs*, 43, 175–223.
- Goldschmidt, Z., & Nissan-Rozen, I. (2020). The intrinsic value of risky prospects. *Synthese*, 198(8).
- Groll, D. (2012). Paternalism, respect, and the will. *Ethics*, 122(4), 692–720. <https://doi.org/10.1086/666500>
- Hausman, D., & McPherson, M. (2009). Preference satisfaction and welfare economics. *Economics and Philosophy*, 25, 1–25. <https://doi.org/10.1017/s0266267108002253>
- Hayenhjelm, M. (2006). Out of the ashes: Hope and vulnerability as explanatory factors in individual risk taking. *Journal of Risk Research*, 9(3), 189–204. <https://doi.org/10.1080/13669870500419537>
- Heath, J. (2020). *The Machinery of Government, chapter Cost-Benefit Analysis as an Expression of Liberal Neutrality*. Oxford University Press.
- Hecht, L. (2021). Permissible risk-inclination for benevolent decision-makers. Unpublished manuscript.
- James, A. (2012). Contractualism's (not so) slippery slope. *Legal Theory*, 18(3), 263–292. <https://doi.org/10.1017/s1352325212100002x>
- Jeffrey, R. (1965/1983). *The Logic of Decision* (2nd ed.). University of Chicago Press.
- Jonker, J. (2020). Risk and asymmetry in development ethics. *African Journal of Business Ethics*, 14(1), 23–41. <https://doi.org/10.15249/14-1-244>
- Mill, J. S. (1859). *On Liberty*. Cambridge University Press.2001.
- Murray, D., & Buchak, L. (2019). Risk and motivation: When the will is required to determine what to do. *Philosophers' Imprint*, 19(16), 1–12.
- Nebel, J. (2020). Rank-weighted utilitarianism and the veil of ignorance. *Ethics*, 131(1), 87–106. <https://doi.org/10.1086/709140>
- Nebel, J., & Buchak, L. (2021). The value of risky prospects. Unpublished handout.
- Nissan-Rozen, I. (2020). Weighing and aggregating reasons under uncertainty: a trilemma. *Philosophical Studies*, 178(9), 2853–2871. Online first. <https://doi.org/10.1007/s11098-020-01587-9>
- Nissan-Rozen, I., & Fiat, J. (2021). Attitudes to risk when choosing for others. Unpublished manuscript.
- Okasha, S. (2016). On the interpretation of decision theory. *Economics and Philosophy*, 32(3), 409–433. <https://doi.org/10.1017/s0266267115000346>
- Otsuka, M. (2015). Prioritarianism and the measure of utility. *The Journal of Political Philosophy*, 23(1), 1–22. <https://doi.org/10.1111/jopp.12023>
- Otsuka, M., & Voorhoeve, A. (2009). Why it matters that some are worse off than others: An argument against the priority view. *Philosophy and Public Affairs*, 37(2), 171–199. <https://doi.org/10.1111/j.1088-4963.2009.01154.x>
- Pettigrew, R. (2015). Risk, rationality, and expected utility theory. *Canadian Journal of Philosophy*, 45(5–6), 798–826. <https://doi.org/10.1080/00455091.2015.1119610>
- Polman, E., & Wu, K. (2020). Decision making for others involving risk: A review and meta-analysis. *Journal of Economic Psychology*, 77(C), 102184. <https://doi.org/10.1016/j.joep.2019.06.007>
- Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior & Organization*, 3(4), 323–343. [https://doi.org/10.1016/0167-2681\(82\)90008-7](https://doi.org/10.1016/0167-2681(82)90008-7)
- Savage, L. (1972). *The Foundations of Statistics* (second revised edition). Dover Publications.
- Scharding, T. K. (2021). Contractualism and risk preferences. *Economics and Philosophy*, 37(2), 260–283. <https://doi.org/10.1017/s0266267120000310>
- Stefansson, H. O. (2021). Ambiguity aversion behind the veil of ignorance. *Synthese*, 198(7), 6159–6182. <https://doi.org/10.1007/s11229-019-02455-8>
- Stefansson, H. O., & Bradley, R. (2019). What is risk aversion? *The British Journal for the Philosophy of Science*, 70(1), 77–102. <https://doi.org/10.1093/bjps/axx035>
- Sugden, R. (2018). *The Community of Advantage: A Behavioural Economist's Defence of the Market*. Oxford University Press.
- Thoma, J. (2019a). Decision theory. In R. Pettigrew & J. Weisberg (Eds.), *The Open Handbook of Formal Epistemology*. PhilPapers.
- Thoma, J. (2019b). Risk aversion and the long run. *Ethics*, 129(2), 230–253. <https://doi.org/10.1086/699256>
- Thoma, J., & Weisberg, J. (2017). Risk writ large. *Philosophical Studies*, 174(9), 2369–2384. <https://doi.org/10.1007/s11098-017-0916-3>
- Vellemann, D. (1993/2000). The story of rational action. In *The Possibility of Practical Reason*. Oxford University Press.
- von Neumann, J., & Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton University Press.
- Weinstein, M. C., Torrance, G., & McGuire, A. (2009). QALYs: The basics. *Value in Health*, 12(s1), S5–S9.
- Weirich, P. (2020). *Rational Responses to Risk*. Oxford University Press.
- Zhao, M. (forthcoming). Ignore risk; maximize expected moral value. *Nous*. <https://doi.org/10.1111/nous.12398>

AUTHOR BIOGRAPHY

Johanna Thoma is Associate Professor of Philosophy at the Department of Philosophy, Logic and Scientific Method at the London School of Economics and Political Science. She works on practical rationality and decision theory, ethics and public policy, and economic methodology.

How to cite this article: Thoma, J. (2023). Taking Risks on Behalf of Another. *Philosophy Compass*, e12898. <https://doi.org/10.1111/phc3.12898>