Economics & Philosophy (2022), **38**, 450–465
doi:10.1017/S0266267121000225

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

A new puzzle in the social evaluation of risk

Marc Fleurbaey[1] and Stéphane Zuber[1,2]

[1]Paris School of Economics and CNRS, Paris, France and [2]University Paris 1, Paris, France

(Received 25 August 2020; revised 27 May 2021; accepted 08 June 2021; first published online 2 February 2022)

Abstract

We highlight a new paradox for the social evaluation of risk that bears on the evaluation of individual well-being rather than social welfare, but has serious implications for social evaluation. The paradox consists in a tension between rationality, respect for individual preferences, and a principle of informational parsimony that excludes individual risk attitudes from the assessment of riskless situations. No evaluation criterion can satisfy these three principles. This impossibility result has implications for the evaluation of social welfare under risk, especially when the preferences of some individuals are not known. It generalizes existing impossibility results, while relying on very weak principles of social rationality and respect for individual preferences. We explore the possibilities opened by weakening each of our three principles and discuss the advantages and drawbacks of these different routes.

Keywords: Social welfare; risk; social rationality; Pareto; fairness

1. Introduction

Harsanyi's (1955) theorem states that, in the context of risk, if social evaluation and individual preferences all take the form of an expected utility, and social evaluation respects individual preferences by satisfying a version of the Pareto principle, then social evaluation must take the form of a weighted sum of the individuals' von Neumann-Morgenstern (VNM) utilities. While Harsanyi and many scholars celebrated this theorem as vindicating utilitarianism, others have argued that one could relax the expected utility assumption for social welfare in order to accommodate considerations of *ex ante* fairness (Diamond 1967), or that one could weaken the Pareto principle in order to accommodate considerations of *ex post* fairness (Fleurbaey 2010). Ultimately, one can read this literature as revealing a tension between social rationality, respect for individual preferences, and fairness. In this literature, fairness is defined as aversion to inequality of some sort (either over individual prospects or over final consumption levels). This aversion to inequality may differ from what is implied by utilitarianism: under utilitarianism, there is no aversion to inequalities in individual expected utilities, whereas inequality aversion over final consumption levels is exactly determined by individuals' aversion to risk (in the sense of Arrow and Pratt).

In this paper, we present an even more fundamental tension, which bears on the assessment of individual well-being rather than social welfare. It is a tension between rationality, respect for individual preferences, and a principle of informational parsimony that excludes individual risk attitudes from the assessment of riskless situations. The principles of rationality and respect for preference principles that we adopt in this paper are fairly intuitive and compelling: in particular, they are milder than other requirements in the literature and avoid objections of spurious unanimity (Mongin 2016). The principle of informational parsimony is less familiar and we need to justify why it is also worth considering.

The principle of informational parsimony states that, in situations not involving risk, we do not need to know people's attitudes to risk in order to compare individual well-being levels. Harsanyi's utilitarianism relies on individuals' Von Neumann Morgenstern (henceforth VNM) utilities, and thus on attitudes towards risk, for interpersonal comparisons. Many applied allocation problems, however, are studied without incorporating risk in the analysis. Sometimes analysts (e.g. Stern 2014) invoke risk attitudes to calibrate the utility functions that are used, even if there is no explicit risk in the contemplated problem. But most often, risk-free problems are studied on the sole basis of preferences over the dimensions at hand, if only because preferences over extra dimensions are not observable in the context that is studied. Many evaluation criteria in cost-benefit analysis and in fair allocation theory rely solely on individual ordinal preferences over commodities and do not bother to examine individual risk attitudes. Our paper shows that either these evaluation criteria are deeply flawed, or they cannot be extended to risky contexts without clashing with compelling principles of rationality and respect for preferences.

Although our main argument is directed towards the evaluation (and comparison) of individual well-being, it has serious implications for the evaluation of collective risk. Indeed, such evaluation generally relies on the evaluation of the situation of the individuals involved in the problem, at least if some unanimity principle like the Pareto principle holds. Our paper thus also contributes to the literature on the aggregation of preferences under risk and uncertainty and presents a new impossibility result in that context. This impossibility result is more fundamental than the readings of Harsanyi's theorem as an impossibility for egalitarians and prioritarians, because it involves a much weaker Pareto condition which, up to now, was considered compatible with fairness principles.

The paper is structured as follows. In section 2 we introduce and prove the impossibility result for individual evaluation. In section 3 we discuss the various options that are available when one of the conditions of the result is relaxed. Section 4 examines how this problem bears on the evaluation of social, as opposed to individual, situations. Section 5 contains a discussion and a tentative conclusion.

2. The impossibility

Consider a single individual who is facing either sure options x in a set X or acts a that map any states of the world s from a finite set S into the set X. Acts are the

https://doi.org/10.1017/S0266267121000225 Published online by Cambridge University Press

standard way of representing risk in the Savage (1972) approach, and we assume that there are only finitely many states for simplicity. For each state s and for an act a, $a(s)$ is the option obtained in state s when choosing a: it is therefore an element of X. For simplicity, we assume that each state s has a well-defined probability.[1] We assume that the individual's preferences over X and over A can be represented with the help of a VNM utility function u. The individual thinks that x is at least as good as x' if and only if $u(x) \geq u(x')$, and that a is at least as good as a' if and only if $Eu(a) \geq Eu(a')$, where Eu denotes the expected utility.

We do not actually need to assume that the individuals herself is an expected utility maximizer. We only posit that the evaluation of risky individual prospects relies on expected utility, and this can be a normative stance for the evaluator rather than an empirical assumption about how the individual behaves. When we talk about respecting individuals' risk attitudes in this paper, this is more fundamentally about respecting the best evaluation of their interests, not necessarily about respecting their possibly flawed attitudes. For instance, they may be ignorant of the probabilities and exhibit ambiguity aversion, or display loss aversion and status quo bias. The best evaluation of their interests is supposed to eliminate such behavioral phenomena and take the most rational approach.[2]

The problem we consider is to assess the well-being of the individual in situations described by an option (a sure one, generally denoted x, or a risky prospect, generally denoted a) and a VNM utility function, generally denoted u. We want to allow for the possibility that the individual may have different utility functions. Such a possibility would be essential to make interpersonal comparisons,[3] but it is also important for a single individual, for instance if we think of assessing the situation of a future individual whose preferences are yet unknown. The concrete question that motivated our research on this problem is the question of sustainability when we do not know whether future generations will care about the environment more than the present generation or will adapt to the degraded environment that we prepare for them.

Formally, we seek an ordering $\succcurlyeq$ over situations (x, u) and (a, u), treating x as a specific type of act that yields a sure outcome. The domain of such situations must be large enough to contain two acts that have the same expected utility with one VNM utility function, but different expected utilities with some other utility function. Hereafter, the notation $(a, u) \succcurlyeq (a', u')$ means that an individual with VNM utility function u facing an act a is better-off than an individual with VNM utility

[1]If we consider instead a situation of uncertainty where probabilities are not well-defined, individuals may have different beliefs, and an impossibility may already arise on the sole basis of the Pareto principle: see Mongin (1995). To focus on our main argument and contribution here, we do not consider this issue in this paper.

[2]Our main argument would however carry over to situations where individuals' best interest itself may even not be represented by an expected utility. The expected utility model is convenient because it identifies a specific utility function u that represents a person's risk attitudes. The description of risk attitudes beyond the expected utility model is more complicated (see for instance Chateauneuf *et al.* (2005) for the rank-dependent expected utility model).

[3]This paper thus follows the idea by Fleurbaey and Tadenuma (2014) of developing "Universal social orderings" for inter-personal and inter-society comparisons.

https://doi.org/10.1017/S0266267121000225 Published online by Cambridge University Press

function u' facing an act a'.[4] This type of comparison could be understood either as an interpersonal comparison of wellbeing from the social point of view, or as an intrapersonal exercise although, for clarity of exposition, this will not be our favored interpretation. Here are the three principles we would like the ordering to satisfy.

The first principle is a very basic rationality principle in the presence of risk. It stipulates that if an option is better than another in all possible states, then it is better. This is a minimal rationality principle, it just imposes consistency between how the ordering assesses prospects *ex ante* and how it assesses final consequences *ex post*.

Statewise dominance: For all a, a', u, u' : [5]
if $(a(s),\ u) \succcurlyeq (a'(s),\ u')$ for all s, then $(a,\ u) \succcurlyeq (a', u')$.

The second principle is that for a fixed utility function, one should respect the individual's preferences over acts. Since there is only one individual under consideration at a time, the objections against the Pareto principle that invoke inequalities do not apply here. We will discuss how social considerations, in particular inequalities, can be taken into consideration in section 4 of this paper.

Pareto: For all a, a', u :
$(a,\ u) \succcurlyeq (a', u)$ if and only if $Eu(a) \geq Eu(a')$.

This Pareto principle is related to the "acceptance principle" put forward by Harsanyi in his impartial observer's argument for utilitarianism (Harsanyi 1977). The acceptance principle applies to situations where the impartial observer knows which identity she will assume for sure. It requires that, in this case, the impartial observer's preferences must coincide with that individual's preferences that are represented by the individual's expected utility.

Observe that the Pareto principle does not imply that *ex ante* evaluations always rely on an expected value of *ex post* consequences (or, said differently, that the social criterion satisfies the VNM independence principle). The Pareto principle implies that social comparisons take the expected utility form only for comparisons of acts involving the same individual but not when we compare two individuals. In section 3, we provide an example of social comparisons satisfying the Pareto principle but *not* Statewise dominance (and therefore they are not of the expected utility type).

The last principle says that only preferences over X matter for the assessment of sure options. In other words, risk attitudes are irrelevant in this context. Formally, this is expressed by considering utility functions that have the same preference ordering over X. We thus introduce the following terminology: we say that two utility functions u and u' are *ordinally equivalent* on X if for all x, x' in X, $u(x) \geq u(x')$ if and only $u'(x) \geq u'(x')$. When u and u' are *ordinally equivalent*

[4]This type of comparisons across utility functions is commonplace (though sometimes implicit in the evaluation of social distributions) in the literature on interpersonal comparisons, as well as the related literature on extended preferences (the latter generally refers to an impartial observer considering the possibility of adopting the preferences of anyone in society). See Adler (2019) for an overview.

[5]In this paper, the notation u' does not refer to the derivative of function u.

https://doi.org/10.1017/S0266267121000225 Published online by Cambridge University Press

on X, they yield the same preference ordering over X. Then, the next principle states that they yield equivalent situations for sure options (the symbol $\sim$ denotes the "equally good" relation generated by $\succcurlyeq$).

Irrelevance of risk attitudes in absence of risk:[6] For all u, u' :

If u and u' are *ordinally equivalent* on X, then for all x in X,

$$(x, u) \sim (x, u').$$

The above principle is related to a widely-used principle in the theory on fair allocation of goods, namely the principle of equal treatment of equals (see Thomson, 2011). Equal treatment of equals stipulates that two individuals with the same preferences should be treated in the same way (and thus should have the same well-being level at a given allocation). Irrelevance of risk attitudes in absence of risk adds the condition that the preferences considered in that case are only those on sure options. This is consistent with our observation, made in the introduction, that risk-free problems are usually studied on the sole basis of preferences over the risk-free outcomes, without any considerations of the risk preferences. A similar line of argument has appeared in Chambers and Echenique (2012).

Irrelevance of risk attitudes in absence of risk can also be related to the idea, put forward by Moreno-Ternero and Roemer (2008), that priority should be given to individuals on the basis of some objective notion of wellbeing when there is no risk. Irrelevance of risk attitudes in absence of risk also implies that we have some notion of wellbeing for non-risky outcomes that is independent of the risk attitudes as represented by the VNM utility function. In our case, contrary to Ternero and Roemer (2008), this wellbeing measure may depend on preferences over non-risky outcomes and sure options.

We can now state our impossibility result.

Theorem: There is no ordering that satisfies Statewise dominance, Pareto and Irrelevance of risk attitudes in absence of risk.

Proof: The proof is very simple. Consider a pair of acts a, a', and two VNM functions u, u' that represent the same preferences over X but not the same risk attitudes, and such that $Eu(a) = Eu(a')$ but $Eu'(a) \neq Eu'(a')$. By Pareto, one has $(a, u) \sim (a', u)$ and not $(a, u') \sim (a', u')$. This reflects the fact that expected utility preferences using two VNM utility functions that are non-affine increasing transformations of one another are different. The contradiction comes, as shown below, from the fact that Statewise dominance and Irrelevance of risk attitudes in absence of risk imply that those preferences should be the same.

Pick a state s. By irrelevance of risk attitudes in absence of risk, and the fact that u and u' are ordinally equivalent on X, $(a(s), u) \sim (a(s), u')$. Since this is true for every s, by

[6]An alternative, logically equivalent, formulation would read: For all u, u', v, v', if u and u' are *ordinally equivalent* on X, and v and v' are *ordinally equivalent* on X, then for all x, x' X, if $(x, u) \succcurlyeq (x', v)$ in and only if $(x, u') \succcurlyeq (x', v')$.

https://doi.org/10.1017/S0266267121000225 Published online by Cambridge University Press

statewise dominance one has $(a, u) \sim (a, u')$. By the same reasoning, $(a', u) \sim (a', u')$. This shows that risk attitudes are irrelevant even in the context of risk.

Recall that $(a, u) \sim (a', u)$. By transitivity, one then has $(a, u') \sim (a', u')$, which yields a contradiction. QED.

This proof shows that one does not need the full force of the Pareto condition to obtain the impossibility. It would be sufficient to posit that risk attitudes do make a difference in at least one case, since the bulk of the proof consists in showing that risk attitudes never matter. Likewise, the theorem is valid if a' is set equal to a and $\succcurlyeq$ is replaced by $\sim$ in the statement of Statewise Dominance.

The logic of the argument is that after uncertainty has been resolved, one should use the evaluation that one would use in absence of risk. But if such evaluation does not rely on risk attitudes, then statewise dominance implies that risk attitudes never matter. What is not intuitive here is that it seems that it should be possible to invoke risk attitudes at least when trading-off the possible outcomes in different states. But the evaluation bears on pairs (act, utility function). Statewise dominance implies that if changing the individual's risk attitude in such a pair never makes a difference in any state of the world, it cannot make a difference at all.

What is surprising about this result is that there are ways to combine Pareto, Irrelevance of risk attitudes in absence of risk and Statewise dominance when preferences are fixed, as showed by Fleurbaey and Zuber (2017) in a context with several individuals (see also section 5). Therefore, the three principles are not obviously incompatible, when doing social choice over prospects for a population with any *given* profile of preferences and risk attitudes. They become incompatible when we seek comparisons of situations involving *different* profiles. This observation appears simple when one understands it, indeed, but it questions the whole literature that has built interpersonal comparisons without taking account of risk attitudes, for riskless allocations.

The next section explores the options that become available when one of the three conditions is dropped.

3. Possibilities

Let us first examine what is possible when statewise dominance is dropped. In fact, there is a very precise possibility, because the other two conditions leave only one class of criteria possible. Rather than stating a characterization theorem, which would involve more notations and formalism, let us explain the reasoning. First, irrelevance of risk attitudes in absence of risk implies that the evaluation of situations (x, u) does not depend on risk attitudes. This is given by an ordering $\widehat{227D}$ such that $(x, u) \widetilde{\widehat{227D}} (x, u')$ whenever u and u' are ordinally equivalent on X and, in order to satisfy Pareto, $(x, u) \widehat{227D} (x', u)$ whenever $u(x) \geq u(x')$.

On the other hand, under some basic continuity assumptions that will not be spelled out here in detail, for every risky situation (a, u), there is at least one sure option $x(a, u)$ that is equally good for the individual, i.e., such that $u(x(a, u)) = Eu(a)$. Let us call this $x(a, u)$ a "certainty-equivalent" of the situation (a, u). The Pareto condition implies that $(x(a, u), u) \sim (a, u)$.

https://doi.org/10.1017/S0266267121000225 Published online by Cambridge University Press

Therefore, only one type of evaluation is possible: for every risky situation, find a certainty-equivalent, and evaluate the certainty-equivalent situations with a criterion that is independent of risk attitudes.[7] This defines a class of orderings, and in this class the various orderings differ by the criterion that is used for riskless situations. But once the criterion for riskless situations is chosen, the evaluation of risky situations is fully determined.

Let us note that, although it violates statewise dominance in general, this approach is consistent with statewise dominance when we consider situations with exactly the same preferences. Indeed, in that case a situation will be better if the certainty equivalent gives higher utility (in terms of the VNM utility function), which is equivalent to the expected utility being higher – but this is only true because the two situations have the same VNM utility function.

What can go wrong with this approach? After all, certainty-equivalence is not uncommon in welfare economics, and there is for instance a popular criterion, the Epstein–Zin (1989) utility, that makes heavy use of it for the assessment of acts unfolding over time in a sequential way. What is bizarre about this approach, in our setting, is the following. Consider u that is more risk-averse than u',[8] and an act a that gives a better outcome in every state of the world than a'. Since u is more risk-averse than u', it is possible that the certainty-equivalent of (a, u) is worse than the certainty-equivalent of (a', u'). Is it really true, then, that (a, u) is worse than (a', u') in spite of its guaranteeing a better outcome? This may not seem outlandish in view of the greater risk aversion with u, but this violation of statewise dominance is bound to raise thorny issues when evaluating well-being along a lifetime. If the certainty-equivalent is computed successively at different periods, time inconsistent evaluations are likely to occur. If, in order to avoid such problems, one wants to only refer to the certainty-equivalent computed at a single period, this opens a wedge between the evaluation and the available information at the time of evaluation.

Let us illustrate this problem with a simple case where options are income levels and there are two equally likely states of the world. One can then assume that VNM functions are increasing functions of income, so that all VNM functions are ordinally equivalent (they all have the same ordering of income levels: more income is always better). Consider Table 1 describing the incomes produced by two acts act a and a' in the two states of the world:

Clearly, with the information that we may have after the state of the world is known, the options produced by act a are better: in state 1 and state 2 it

[7]Formally, the criteria would be as follows: for all a, a', u and u', $(a, u') \succcurlyeq (a', u')$ if and only if $(x(a, u), u)\widehat{227}\mathrm{D}(x(a', u'), u')$. When $u = u'$, the condition boils down to $u(x(a, u)) \geq u(x(a', u))$, which by definition of the certainty-equivalent can be written $Eu(a) \geq Eu(a')$. The Pareto principle is satisfied. When $a = a' = x$, we have $x(a, u) = x(a', u') = x$ and thus the condition is equivalent to $(x, u)\widehat{227}\mathrm{D}(x, u')$.

[8]Risk-aversion is defined here in the sense of Arrow (1971) and Pratt (1964), and generalized to multidimensional outcomes by Kihlstrom and Mirman (1974). These definitions hold for the expected utility model. The literature in philosophy often defines risk-aversion through a non-expected utility model, where probabilities are being transformed in some way (see Buchak 2013). Some people argue that we should talk about risk aversion only in the second case (Hansson 1988). In this paper, we adopt the more general view that risk-aversion can be defined more generally with respect to any relevant metric (in our example, this metric is money).

https://doi.org/10.1017/S0266267121000225 Published online by Cambridge University Press

Table 1. Statewise dominance vs. certainty-equivalent

	State 1	State 2
Act a	10000	40000
Act a'	9000	37000

produced more income, which is the only relevant dimension for social evaluation (we assume that individuals are the same in all respects: they have the same needs and none of them deserves more). It may thus appear appealing to prefer act a. But this is not necessarily the case if we adopt the criterion based on the certainty-equivalent. Assume for instance that u is such that $u(x) = x^{1/2}$ for all income levels x, while $u'(x) = x$. It can be showed that the certainty equivalent of (a, u) corresponds to an income of 22500, while the certainty-equivalent of (a', u') is 23000. Given that the only criterion we have that is independent of risk attitudes is to compare income levels, we should thus prefer (a', u') based on the information we have before the state of the world is known. This inconsistency between the social evaluation before or after the state of the world is known, which is due to a violation of statewise dominance, seriously undermines the credibility of any evaluation that would proceed by dropping state-wise dominance.

To be more concrete, let us consider the case of an education programme that can reduce the risk aversion of future people. This programme has a cost that will be borne by the future people themselves. This cost makes them ultimately worse off, and we know this for sure *ex ante*. The programme may raise their certainty-equivalent, but it seems odd to want to implement the programme.

Violations of Statewise dominance can trigger violations of time consistency, because the *ex ante* evaluation (not knowing the state of the world) may induce planning certain future actions which, *ex post* (knowing the state of the world), appear dominated. Suppose the education program taken as an example in the previous paragraph is planned for the next generation because it enhances its certainty-equivalent. Once more information about the situation of the next generation is made available, however, it appears harmful to pursue the program as planned and the program would better be cancelled. The key rationality problem with a violation of time consistency is that it is known in advance that the chosen plan will be cancelled later in every state of the world. Why, then, plan a policy that will be repealed for sure? It thus seems to us that dropping Statewise dominance is probably the least appealing route: its seems preferable to consider dropping either Pareto or Irrelevance of risk attitudes in absence of risk.

Let us first consider the Pareto principle. Instead of completely abandoning the Pareto principle, we can try to weaken it and retain Pareto only for riskless options (x, u). This is the natural option since for the impossibility to arise, it suffices to have a pair of acts with a ranking that depends on risk attitudes. The VNM utility function u is a particular cardinal representation of preferences over riskless options, and we can choose another representation for each u that satisfies two conditions: (i) u and are ordinally equivalent on X; (ii) whenever u, u' are ordinally equivalent on X, one has $\overline{u} = \overline{u'}$. In other words, the transformation respects the preferences over X

https://doi.org/10.1017/S0266267121000225 Published online by Cambridge University Press

but eliminates differences in risk attitudes by changing the cardinal scale of utility measurement. Consider the ordering defined by: $(a, u) \succcurlyeq (a', u')$ if and only if $E\overline{u}(a) \geq E\overline{u'}(a')$. This criterion satisfies statewise dominance because it relies on expected utility, and it satisfies irrelevance of risk attitudes in absence of risk by construction.

The example above is just one example, but to get the entire class of solutions that are available with the weakening of Pareto to riskless options, one only has to include all decision criteria under risk that satisfy statewise dominance but do not take the form of an expected utility.

Nevertheless, the fact that this approach does not respect risk attitudes in general may make it evaluate risky prospects in a starkly different way than the individual, or what is in her best interest. Unless there is a reason not to trust risk attitudes, this seems quite disrespectful. Imagine for instance that the normalized utilities are risk neutral in some relevant sense whereas the individual is very risk averse. Such an approach would then vastly overestimate the well-being of the individual when facing very high risks with positive expected payoffs but substantial probability of a high loss.

Finally, the third possibility consists in dropping Irrelevance of risk attitudes in absence of risk. The simplest criterion is then to rely simply on $Eu(a)$ for the evaluation of (a, u), but this may be simplistic, because the VNM functions are defined by risk attitudes only up to an affine transform. Therefore, it is probably more interesting to think that, again, a normalization is in order, though one that respects not only preferences on X but also risk attitudes. Concretely, this means that, for each utility function u, there would be two numbers $\alpha_u > 0$ and β_u, such that: $(a, u) \succcurlyeq (a', u')$ if and only if $E[\alpha_u u(a) + \beta_u] \geq E[\alpha_{u'} u'(a') + \beta_{u'}(u')]$. An analysis of the normalization of VNM utilities in applications of utilitarianism is made in Fleurbaey and Zuber (2018).

The approach defined in the previous paragraph would be the only acceptable one if statewise dominance was strengthened into requiring that the ordering $\succcurlyeq$ be itself representable by an expected utility. Since it seems reasonable to us that expected utility is the most rational criterion under risk, we consider that this approach is therefore the one that deserves most attention when irrelevance of risk attitudes in absence of risk is dropped.

One may wonder if, like Pareto, irrelevance of risk attitudes in absence of risk could be weakened rather than completely dropped. But for riskless options, the formula $\alpha_u u(x) + \beta_u$ depends on the curvature of u in a crucial way and it appears impossible to define an interesting subset of cases in which risk attitudes would not influence the evaluation.

4. Implications for social evaluation

The impossibility highlighted in this paper is really about the evaluation of individual situations, but it undermines the construction of social preferences, because social preferences rely on interpersonal comparisons as an essential building block.

https://doi.org/10.1017/S0266267121000225 Published online by Cambridge University Press

One can also reformulate the impossibility theorem in terms of the evaluation of social risk based on some social preference ordering. Let us thus seek an ordering, still denoted $\succcurlyeq$, over social situations, which are either sure situations $(x, u_i)_{i\in P}$ or acts $(a, u_i)_{i\in P}$, where i is the index denoting an individual and P is the population.

The axiom of Statewise dominance can be kept unchanged, except for notation details. The axiom of Irrelevance of risk attitudes in absence of risk can also be kept, specifying that each individual's VNM can change without affecting the social ordering on sure situations, provided individual preferences over sure situations remain unchanged.

The Pareto axiom is more problematic. When should we respect individual risk attitudes when there are several individuals and the risky prospect may generate different levels of inequality? Fleurbaey (2018) reviews several axioms proposed in the literature and concludes that the least controversial consists in restricting the application of the Pareto principle to situations which are as close as possible to single-individual situations. Such situations are characterized by the absence of inequalities in every state of the world. If there is risk, everyone is then "in the same boat."

But how should inequalities be measured? Let us assume that there is a given measure of well-being for interpersonal comparisons, a specific utility function v_i for every i. In all generality, this function may or may not be identical to i's VNM utility. Since it is about comparing individuals in final states, one should expect that, in the presence of the axiom of Irrelevance of risk attitudes in absence of risk, the function v_i does not depend on risk attitudes. But we do not make this assumption at the outset. However, we assume that for every u function there is only one v function.

We obtain the following Pareto axiom:

Pareto for equal risk: For all $a, a', (u_i)_{i\in P}$ such that for all s, all $i, j \in P$, $v_i(a(s)) = v_j(a(s))$ and $v_i(a'(s)) = v_j(a'(s))$:

- $(a, u_i)_{i\in P} \succeq (a', u_i)_{i\in P}$ if for all i, $Eu_i(a) \geq Eu_i(a')$;
- $(a, u_i)_{i\in P} \succ (a', u_i)_{i\in P}$ if, in addition, for some i, $Eu_i(a) > Eu_i(a')$.

Observe that this axiom implies the Pareto axiom of the previous sections, when there is only one individual.[9] We now have the following impossibility.

Theorem: There is no ordering that satisfies the three conditions.

Proof: The proof, the details of which are omitted here, mimics the previous proof. Consider two egalitarian acts (acts such that the options obtained in each state of the world are egalitarian, in the sense of the v_i functions). Consider a utility profile $(u_i)_{i\in P}$ where all individuals have the same VNM function, and another utility profile $(u_i')_{i\in P}$ where all individuals have the same VNM function, which is ordinally equivalent to that in the first profile but involves different risk attitudes. Assume

[9]Our result also holds true for other Pareto conditions, such as one applying when only one individual is submitted to risk and other individuals bear no risk and are indifferent between the two options, or the stronger one applying when individuals are submitted to independent risks (see Bommier and Zuber 2008 for use of such a Pareto condition).

https://doi.org/10.1017/S0266267121000225 Published online by Cambridge University Press

that the two acts are indifferent with the first utility profile, but that one act is preferred to the other with the second profile. By Pareto for equal risk, the social preference is just the same as the individual preference.

By Irrelevance of risk attitudes in absence of risk and Statewise dominance, once again, risk attitudes do not matter, which contradicts the application of the Pareto principle in the previous paragraph. QED.

As we did for the evaluation of individual risk, we can try to look for possibilities by weakening one of the axioms. Salient solutions, similar to those discussed in section 3, are as follows:

a) If we drop Statewise dominance, we can focus on the distribution of certainty-equivalent situations over the population. For instance, if options are just income levels, one could adopt an additive separable social welfare function that would consist in the sum of some concave transformation of certainty-equivalent incomes. One then obtains a version of *ex ante* prioritarianism as discussed by Adler and Treich (2017).
b) If we modify Pareto by restricting it to riskless (but not necessarily equal) situations, we can measure individual well-being by the expected value of some normalized utility function that normalizes risk attitudes among people having ordinally equivalent utilities over sure options. Then taking a sum of these individual well-being measures provides a social welfare function that satisfies Statewise dominance (actually the stronger expected utility requirement), Irrelevance of risk attitudes in absence of risk and the modified Pareto requirement.
c) If we drop Irrelevance of risk attitudes in absence of risk, we are left with Pareto for equal risk and Statewise dominance. Social orderings satisfying these two requirements are extensively studied in Fleurbaey and Zuber (2017) for the case of fixed profiles of preferences. If Statewise dominance is strengthened to an Expected utility requirement for the social ordering, Fleurbaey and Zuber (2017, Theorem 1) show that we obtain a specific class of criteria that work in two steps. First, in each state of the world we compute an equally distributed equivalent option, that is, an egalitarian option (in the sense of the v_i functions) that is equivalent from the social point of view. Then we take an affine aggregation of individual expected utility at this egalitarian prospect, using the actual VNM functions of individuals. The normalization of the VNM functions for interpersonal comparisons and cross-profile assessments can then be determined as discussed in Adler (2019) or Fleurbaey and Zuber (2018).

5. Discussion and conclusion

The conundrum raised by our impossibility theorem relates to two strands of the literature.

https://doi.org/10.1017/S0266267121000225 Published online by Cambridge University Press

As already mentioned in the introduction, Harsanyi's theorem (Harsanyi, 1955) reveals a tension between social rationality, Pareto and fairness (both *ex ante* and *ex post*). In this tension, fairness is meant to refer to inequality aversion either with respect to individual expected utilities (*ex ante* fairness) or to final well-being (*ex post* fairness). Our axiom of irrelevance of risk attitudes in absence of risk is not about inequality aversion at all, but it does relate to fairness indirectly. As discussed before, in the theory of fair allocation, interpersonal comparisons are typically made in terms of bundles of commodities and ordinal preferences, and therefore such comparisons ignore risk attitudes when there is no risk in the contemplated allocation problem. Fleurbaey and Zuber's (2017) variant of Harsanyi's result mentioned as the end of the previous section applies the weighted utilitarian sum of utilities not to individual bundles, but to hypothetical final allocations in which inequalities have been eliminated, in every state of the world, while preserving social welfare as defined for riskless allocations, i.e. ignoring risk attitudes. This offers a nice combination of the standard theory of fair allocation for riskless settings and Harsanyi's theorem suitably watered down. But it was proposed for a situation in which the composition of the population and individual preferences are fixed. This reconciliation is seriously undermined by our result here, in the perspective of dealing with preference changes and changes in the composition of the population.

The other strand of the literature to which our result relates has been looking at the tension between dominance principles and context-dependence principles. The main context in which this tension has been studied (Fleurbaey 2007; Pattanaik and Xu 2007) involves interpersonal comparisons of bundles of commodities where the context-dependence idea is represented by a sort of Pareto principle, in the following form: when preferences change, comparisons of bundles are altered (i.e., whether bundle x is better than bundle y depends on the preferences of the individual enjoying them). In the present context, the relevant dominance principle is bundle dominance: if one individual is better off in every dimension, this individual is better off overall.[10]

The bundle dominance principle is endorsed in the literature in various contexts. In particular, Sen (1992) defends it as a cornerstone of the evaluation of capabilities. The tension between this principle and Pareto comes from the fact that for certain preferences, bundles x and y may be equally good, whereas for a second type of preferences, bundles x^+ and y^- may be equally good, and this clashes with the dominance principle if bundle x^+ dominates x while y dominates y^-. This is illustrated in Figure 1, where bundles are bundles of two commodities and the curves are indifference curves for individuals with utility u and u' over bundles.

In Figure 1, the solid line describes all bundles equally good as x with utility function u, while the dotted line describes all bundles equally good as $x^+$with utility function u'. By bundle dominance (x^+, u') is better than (x, u) and (y, u) is better than (y^-, u'), simply because there are more commodities in x^+ than in x, and there are more commodities in y than in y^-. But by Pareto indifference (x, u) is as good as (y, u) (x and y are on the same indifference curve for u), and (x^+, u') is as good as (y^-, u'). Assuming transitivity, we thus obtain an inconsistency in the social evaluation.

[10]Pattanaik and Xu (2012) and Weymark (2017) study the general structure of such dilemmas.

https://doi.org/10.1017/S0266267121000225 Published online by Cambridge University Press

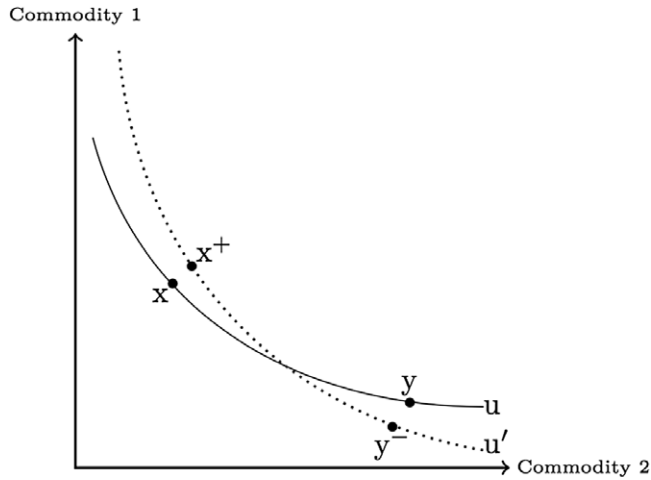

Figure 1. The conflict between bundle dominance and Pareto indifference.

In our paper, we also have the Pareto principle, but bundle dominance is replaced in our analysis by Statewise dominance applied not to physical payoffs but to the evaluation itself, which is a very substantial weakening. As discussed in the previous section, there is no direct clash between Pareto and Statewise dominance, and it is only the third axiom of irrelevance of risk attitudes in absence of risk that triggers a conflict. This third axiom is in fact embedded in the usual bundle dominance principles, in which the statement "more in one dimension" is tacitly assumed to depend only on physical characteristics of the bundle and not on the preferences of the individual over the whole bundle. In our setting, the axiom requires the evaluation of the final situation in every state of the world not to depend on risk attitudes.

Fleurbaey and Blanchet (2013) criticize the bundle-dominance principle for failing to take account of the problem of fit between preferences and bundles. Having more in every dimension may not mean much if associated with different preferences for which the composition of the contemplated bundle is bad. A better paid banker may be worse off than a less well paid colleague if the banker would rather be an artist. We think, however, that this criticism does not apply to our setting. Statewise dominance is a pure rationality principle and has nothing to do with the fit between preferences and prospects. Irrelevance of risk attitudes in absence of risk also cannot be related to the fitting issue either, because it simply says that overall preferences (over prospects) do not matter when risk is absent, i.e. in a particular context.

It is in fact impossible to adapt our formalism to the bundle evaluation problem, because in our setting we can equate a final consequence with a sure situation once uncertainty is resolved, whereas in the bundle context, it would make no sense to imagine living in a world in which only one dimension exists (say, income) and the other dimensions have vanished. There is, however, something similar to the artist-banker example in our setting. This is the case of a risk averse individual who faces a

https://doi.org/10.1017/S0266267121000225 Published online by Cambridge University Press

better prospect but is actually worse-off because he is more risk-averse, so that the risk in the prospect is a "bad fit" for his risk attitude. Supposing that there is something true about this assessment, there are two ways of accommodating it here. One consists in dropping Statewise dominance and evaluating certainty-equivalent situations. A more risk-averse individual can have a lower certainty-equivalent due to risk aversion. But we have seen that violating Statewise dominance leads to practical difficulties because of the induced irrationality in the evaluation.

The other possibility is to drop irrelevance of risk attitudes in absence of risk and posit that a risk-averse individual who ends up being lucky may still have to be considered relatively badly off, compared to an equally lucky but less risk-averse individual, because of risk aversion embodied in the utility function. This means that risk aversion should be incorporated into the evaluation of sure options, and be viewed as a burden even in absence of risk. It then becomes a burden in the presence of risk by the expected value of the burden it has in every possible final consequence. As shown in Fleurbaey and Zuber (2018), it is possible to scale VNM utilities so that more risk averse preferences yield a lower utility for every sure option. In that case, we obtain that VNM utilities, once properly normalized, are (interpersonally comparable) measures of individual well-being.[11]

But one might also want to object that there is no specific burden to be attached to risks borne by such risk averse persons. The fact that some people are more risk averse than others may just be that they rely on different reasons when making their choices in risky situations (for instance thinking in terms of threats rather than thinking in terms of opportunities). If we follow Mongin (2016) in defining spurious unanimity as situations where unanimous preferences are based on mutually incompatible reasons, and if we admit that the Pareto principle has less appeal in situations of spurious unanimity, we may prefer not to rely on risk attitudes for well-being assessment. In that case, we may prefer to retain Statewise dominance and Irrelevance of risk attitudes in absence of risk.

In conclusion, it seems to us that there are two reasonable escape routes from the impossibility presented in this paper. The first one consists in accepting the use of VNM utilities in interpersonal comparisons, even in the absence of risk, and to scale them in a way that makes risk aversion appear as a burden, while retaining expected utility as the valuation measure in order to abide by the most basic principles of rationality and respect for preferences and risk attitudes. The second one consists in grounding social evaluation on the expected value of some utility function that may not correspond to the individual VNM functions but that is consistent with individual ordinal preferences over riskless options. As suggested by the reference to spurious unanimity above, the choice of the appropriate utility function in that case should be based on reasons that are deemed socially acceptable. One possibility would have reference risk attitudes be an aggregation

[11]Adler (2012, 2016, 2019) also defends a measure of individual well-being based on VNM utilities with several arguments based on diverse considerations. Adler's scaling makes more risk-averse individual appear better off between two reference situations. Here we suggest instead that risk aversion is a burden for every situation.

https://doi.org/10.1017/S0266267121000225 Published online by Cambridge University Press

of individual ones in a reference population (for instance taking a weighted sum of individuals' VNM utility functions).

Acknowledgments. This research has been supported by the Agence nationale de la recherche through the Fair-ClimPop project (ANR-16-CE03-0001-01), the "Coping with Heterogeneous Opinions" project (ANR-17-CE26-0003) and the Investissements d'Avenir program (ANR-10-LABX-93), and by the Riksbankens Jubileumsfond (the Swedish Foundation for Humanities and Social Sciences). We are grateful for helpful comments by the editor, two referees, Jean Baccelli, John Broome, Paul Egré, Simon Grant, Philippe Mongin, Klaus Nehring, Wlodek Rabinowicz and Katie Steele, as well as audience at the London School of Economics and the Australian National University.

References

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

Adler M.D. 2016. Extended preferences. In *The Oxford Handbook of Well-Being and Public Policy*, eds M.D. Adler and M. Fleurbaey, 476–517. Oxford: Oxford University Press.

Adler M.D. 2019. *Measuring Social Welfare: An Introduction*. Oxford: Oxford University Press.

Adler M.D. and N. Treich 2017. Utilitarianism, prioritarianism, and intergenerational equity: a cake eating model. *Mathematical Social Sciences* **87**, 94–102.

Arrow K.A. 1971. *Essays in the Theory of Risk Bearing*. Chicago, IL: Markham.

Bommier A. and S. Zuber 2008. Can preferences for catastrophe avoidance reconcile social discounting with intergenerational equity? *Social Choice and Welfare* **31**, 415–434.

Buchak L. 2013. *Risk and Rationality*. Oxford: Oxford University Press.

Chambers C.P. and F. Echenique 2012. When does aggregation reduce risk aversion? *Games and Economic Behavior* **76**, 582–595.

Chateauneuf A., M. Cohen and I. Meilijson 2005. More pessimism than greediness: a characterization of monotone risk aversion in the rank-dependent expected utility model. *Economic Theory* **25**, 649–667.

Diamond P.A. 1967. Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy* **63**, 309–321.

Epstein L.G. and S.E. Zin 1989. Substitution, risk aversion, and the temporal behavior of consumption and asset returns: a theoretical framework. *Econometrica* **57**, 937–969.

Fleurbaey M. 2007. Social choice and the indexing dilemma. *Social Choice and Welfare* **29**, 633–648.

Fleurbaey M. 2010. Assessing risky social situations. *Journal of Political Economy* **118**, 649–680.

Fleurbaey M. 2018. Welfare economics, risk and uncertainty. *Canadian Journal of Economics* **51**, 5–40.

Fleurbaey M. and D. Blanchet 2013. *Beyond GDP*. Oxford: Oxford University Press.

Fleurbaey M. and K. Tadenuma 2014. Universal social orderings: an integrated theory of policy evaluation, inter-society comparisons, and interpersonal comparisons. *Review of Economic Studies* **81**, 1071–1101.

Fleurbaey M. and S. Zuber 2017. Fair management of social risk. *Journal of Economic Theory* **169**, 666–706.

Fleurbaey M. and S. Zuber 2021. Fair utilitarianism. *American Economic Journal: Microeconomics* **13**, 370–401.

Hansson B. 1988. Risk aversion as a problem of conjoint measurement. In *Decision, Probability and Utility*, eds P. Gärdenfors and N.-E. Sahlin, 136–158. Cambridge: Cambridge University Press.

Harsanyi J.C. 1955. Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy* **63**, 309–321.

Harsanyi J.C. 1977. *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*. Cambridge: Cambridge University Press.

Kihlstrom R.E. and L.J. Mirman 1974. Risk aversion with many commodities. *Journal of Economic Theory* **8**, 361–388.

Mongin P. 1995. Consistent Bayesian aggregation. *Journal of Economic Theory* **66**, 313–351.

Mongin P. 2016. Spurious unanimity and the Pareto principle. *Economics and Philosophy* **32**, 511–532.

Moreno-Ternero J. and J.E. Roemer 2008. The veil of ignorance violates priority. *Economics and Philosophy* **24**, 233–257.

Pattanaik P.K. and Y. Xu 2007. Minimal relativism, dominance, and standard of living comparisons based on functionings. *Oxford Economic Papers* **59**, 354–374.

Pattanaik P.K. and Y. Xu 2012. On dominance and context-dependence in decisions involving multiple attributes. *Economics and Philosophy* **28**, 117–132.

Pratt J.W. 1964. Risk aversion in the small and in the large. *Econometrica* **32**, 122–136.

Savage L.J. 1972. *The Foundations of Statistics* (2nd edn). New York, NY: Dover.

Sen A.K. 1992. *Inequality Re-examined*. Oxford: Oxford University Press.

Stern N. 2014. Ethics, equity and the economics of climate change. Paper 2: Economics and politics. *Economics and Philosophy* **30**, 445–501.

Thomson W. 2011. Fair allocation rules. In *Handbook of Social Choice and Welfare*, Volume 2, eds K.J. Arrow, A.K. Sen and K. Suzumura, 393–506. Amsterdam: North-Holland.

Weymark J.A. 2017. Conundrums for nonconsequentialists. *Social Choice and Welfare* **48**, 269–295.

Cite this article: Fleurbaey M and Zuber S (2022). A new puzzle in the social evaluation of risk. *Economics & Philosophy* **38**, 450–465. https://doi.org/10.1017/S0266267121000225