Economics & Philosophy (2026), 1–27
doi:10.1017/S0266267126100716

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

The Risk-Priority View

Pietro Cibinel

Department of Philosophy, University of North Carolina at Chapel Hill, Chapel Hill, USA and Department of Philosophy, Princeton University, Princeton, USA
Email: pcibinel@unc.edu

(Received 07 July 2025; revised 04 December 2025; accepted 03 January 2026)

Abstract

This paper develops and defends a non-utilitarian interpretation of John Harsanyi's social aggregation theorem and sum of vNM utilities approach. On this interpretation, vNM utilities transform an independently available cardinal measure of fully comparable individual well-being. The resulting proposal for ranking well-being distributions – the Risk-Priority View – is not welfare-anonymous and can favour a smaller increase in well-being for one individual rather than a larger increase in well-being for another, equally well-off individual. I argue here that such counterintuitive implications can be defended, and that impartiality can still be secured through the imposition of an alternative, interprofile anonymity axiom.

Keywords: Harsanyi's social aggregation theorem; Sen–Weymark critique; vNM utility; well-being; distributive ethics

1. Introduction

How should we evaluate distributions of well-being across people? The simplest approach is *Utilitarianism*: the value of a distribution is given by summing each person's well-being. Another common approach is the *Priority View*, or *Prioritarianism*: the value of a distribution is given by first transforming each person's well-being in a certain way, and then summing such transformed well-being levels. The prioritarian transformation is the same for all individuals, and ends up giving more weight to worse-off people (Parfit 2002; Adler 2011).

A central argument for Utilitarianism is given by John Harsanyi's interpretation of his social aggregation theorem (1955, 1977, 1982). Harsanyi shows that if individual and social preferences in conditions of risk satisfy the axioms of expected utility theory, and if social preference defers to individual preferences when the latter all agree, then there exist von Neumann–Morgenstern (vNM) expected utility representations $V, V_1, \ldots, V_n$ for society and individual preferences respectively, such that

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (https://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided that no alterations are made and the original article is properly cited. The written permission of Cambridge University Press or the rights holder(s) must be obtained prior to any commercial use and/or adaptation of the article.

https://doi.org/10.1017/S0266267126100716 Published online by Cambridge University Press

$$V = V_1 + \ldots + V_n$$

(following Broome's 2015 statement). Lotteries and hence degenerate lotteries – i.e. *outcomes* – are ranked according to a sum of vNM utilities. Harsanyi went on to interpret the individual vNM utility representations in this formula as cardinal measures of individual well-being. He therefore took his theorem to support Utilitarianism: if outcomes are ranked by a sum of individual vNM utilities, and the latter are cardinal measures of well-being, then outcomes are ranked by summing individual well-being.

However, Harsanyi's interpretation of his social aggregation theorem is contentious. According to what has come to be known as the Sen–Weymark critique (Sen 1976, 1977; Weymark 1991; Roemer 1996), the cardinal structure that individual well-being has (or may have) is to be sharply distinguished conceptually from the quantity that the vNM representation theorem identifies, i.e. which individuals that satisfy the vNM axioms maximize in expectation (von Neumann Morgenstern 1944; see Fishburn 1989). Absent an argument for why the two should be identified, Harsanyi's aggregation theorem fails to support Utilitarianism.

There are different ways one can react to the Sen–Weymark critique. One option is to claim that well-being has no cardinal structure independently of vNM utility; hence there is only really one quantity, vNM utility, and *it* provides a precisification of what we might mean by "quantities of well-being" and "utilitarianism".[1] Another option is to insist that well-being does have cardinal structure independently of vNM utility, and reject the sum of vNM utility approach (more on this below).[2]

This paper explores a different route. Suppose we assume that well-being has cardinal structure independently of vNM utility and is comparable across individuals, and that individuals' preferences can display differing degrees of risk-aversion in well-being, as encoded by their vNM utility.[3] We can then ask two sets of questions. First, what position do we land on, with respect to the question of how *well-being itself* (as opposed to income or resources) should be distributed? That is, how should we then interpret the sum of vNM utilities vis-à-vis the project of ranking well-being distributions? Second, could we still defensibly maintain that social value is given by a sum of vNM utilities? Can the resulting position be morally justified? The large literature surrounding Harsanyi's aggregation theorem seems to lack a sustained philosophical investigation of these questions, especially with regard to the latter set.[4]

[1]See Broome (1991, 2004), Jensen (1995), Risse (2002), McCarthy (2006, 2007) and Greaves (2017) for variants of this approach.

[2]For example, so-called "primitivist" prioritarians (following Greaves' 2015 terminology) adopt this approach. One could also accept the identification of well-being with vNM utility (Adler 2019: 271–2) but reject the resultant utilitarian position in favour of e.g. Prioritarianism (Adler 2011).

[3]Such a cardinal structure could be obtained in one of several ways – e.g. by looking at hedonic intensity (Kahneman *et al.* 1997), strength of preference elicited via conjoint measurement (Bell and Raiffa 1988), metrics of human flourishing (VanderWeele 2017), or the axioms of difference measurement (Krantz *et al.* 1971: Ch. 4; Nebel 2024; Adler 2025). I defend my preferred account in a companion paper (see Cibinel 2025a: Ch. 1). Dietrich (Forthcoming) shows how vNM utility can be decomposed into "pure risk attitude" and "pure utility" or well-being, lending further support to this approach.

[4]For relevant prior work on the former, see Blackorby *et al.* (1980: 19–31; 2008: 159), Broome (1991: 211–213), Blackorby *et al.* (2002, 2005: Ch. 7), Dietrich (Forthcoming), and Cibinel (2025b), in which I introduce the name "Risk-Priority View".

https://doi.org/10.1017/S0266267126100716 Published online by Cambridge University Press

I will treat the Sen–Weymark critique as a point of departure to investigate whether a *non-utilitarian* philosophical defence of the sum of vNM utility approach can be given on the assumption that well-being is cardinally measurable and comparable independently of vNM utility. I develop and defend an alternative interpretation of the philosophical significance of Harsanyi's social aggregation theorem. Like Prioritarianism, the ranking of well-being distributions that I claim is supported by Harsanyi's social aggregation theorem evaluates each distribution of well-being in terms of a sum of transformed well-being levels. Unlike Prioritarianism, it does not perform the same transformation for all individuals, for the simple reason that people's preferences under risk with respect to well-being may differ: what transformation is performed on an individual's well-being level depends on that individual's attitude to risks involving well-being. We can call this a

> **Risk-Prioritarian interpretation of the sum of vNM utilities**. Social value is a sum of transformed well-being levels such that each individual's transformation function is determined by their risk attitude in well-being, with risk attitudes in well-being allowed to vary across individuals.

Is this reinterpretation of the sum of vNM utility approach philosophically viable? An obvious difficulty is that when people's risk attitudes for well-being differ, as they can be expected to, the Risk-Prioritarian interpretation entails that individuals should receive differential treatment *in well-being* (see Blackorby *et al.* 1980: 30; Cibinel 2025b). This contravenes deeply held intuitions. It is standardly believed that

> (Outcome Anonymity) if the well-being distributions of two outcomes x and y only differ in who gets which well-being level, x and y are equally good (see e.g. Adler 2019: 97),

and, even more strongly, that

> (Suppes–Sen Principle) if the well-being distributions of two outcomes x and y only differ in who gets which well-being level, and we improve the well-being of some individuals in x to get $x+$, then $x+$ is better than y (adapting Suppes 1966; Sen 2017: 211).

To think otherwise, it is typically said, amounts to being objectionably partial: to prioritize the interest of some individuals just because of their identity. Utilitarianism, Prioritarianism and all the other main rankings of well-being distributions satisfy these principles. I will argue, however, that the Risk-Prioritarian rejection of Outcome Anonymity and the Suppes–Sen Principle can be defended. The Risk-Priority View denies these principles not because it thinks some individuals' well-being matters more *simpliciter*, but rather because it thinks their well-being matters more in some specific cases, and the well-being of other individuals matters more in other cases. It is a differential treatment that is compatible with, and indeed secures, impartiality.[5]

[5]For a recent examination of the relationship between impartiality and different anonymity principles, see Muñoz (2025).

https://doi.org/10.1017/S0266267126100716 Published online by Cambridge University Press

Here is how the paper will be structured. In section 2, I develop the Risk-Prioritarian interpretation in full detail. In section 3, I discuss the issue of vNM utility comparisons in light of the Risk-Prioritarian interpretation. In section 4, I proceed to develop two interrelated philosophical defences of the Risk-Prioritarian rejection of Outcome Anonymity and the Suppes–Sen Principle. Section 5 concludes the paper, and is followed by a formal Appendix.

Before I proceed, a note on the scope of this paper. My aim is to argue that the Risk-Prioritarian interpretation of the sum of vNM utility approach is philosophically tenable, in the sense that its counterintuitive ranking of well-being distributions can be defended. To my knowledge, no other work in the literature argues for the plausibility of the sum of vNM utility approach in the face of an explicit assumption of cardinal well-being measurability and comparability independent of vNM utility itself, with vNM utility obtained from heterogeneous individual preferences over well-being lotteries. This is the distinctive contribution of the paper. I do *not* attempt to defend the sum of vNM utility approach in general as against a range of objections stemming from alternative approaches one might take in social ethics, and in response to Harsanyi's aggregation theorem. A few such approaches, however, deserve mention. *Ex post* versions of prioritarianism as well as non-separable egalitarianism reject the application of Pareto principles *ex ante*, because of the resulting insensitivity to the distribution of well-being *ex post* (see Adler 2011: Ch. 7; 2025: 241; Voorhoeve and Fleurbaey 2013; Adler and Holtug 2019). *Ex ante* versions of prioritarianism and non-separable egalitarianism, by contrast, reject the application of expected utility theory (indeed, of weaker dominance principles) at the level of social preference; relevant here is Diamond's (1967) early critique of Harsanyi, showing that the sum of vNM utility approach cannot accommodate a kind of "fairness" in the distribution of probabilities of well-being improvements and losses (see Epstein and Segal 1992; McCarthy 2006; for applied comparisons between the *ex ante* and *ex post* prioritarian approaches see Adler *et al.* 2021 and Ferranna *et al.* 2023). A sum of vNM utility proponent certainly needs to reckon with such alternatives, but they can do so independently of how vNM utility is understood: whether as a cardinal measure of well-being, as per the utilitarian interpretation, or as a transformation on well-being, as per the Risk-Prioritarian interpretation. This brief tour of alternatives should, however, underscore how different Risk-Prioritarianism is from traditional strands of prioritarian thought.

2. Setup and Characterization

Well-being is assumed to be cardinally measurable and fully comparable independently of vNM utility.[6] There is a set of n individuals, and a set of final outcomes. A well-being function, $W(\cdot)$, takes in outcomes and returns well-being distributions, understood as lists, with n entries, of numbers representing well-being

[6]The framework I use combines different extant approaches; see, in particular, Blackorby *et al.* (2005: Ch. 7), Blackorby *et al.* (2008), Chambers and Echenique (2012), Fleurbaey and Zuber (2017), Sen (2017), Adler (2019: 271–2), Nebel (2022, 2024) and Cibinel (2025b).

https://doi.org/10.1017/S0266267126100716 Published online by Cambridge University Press

levels on a scale encoding the assumptions about well-being just made: this is known as an *interval scale*, common for all individuals.

We then consider the set of probability functions on the set of outcomes, and interpret such functions as the policies or "lotteries" with possibly uncertain outcomes that we want to evaluate socially.[7] We can now treat any outcome x as equivalent to the "degenerate" probability function that assigns probability 1 to x and 0 to all other outcomes. This allows us to deal with a single object for social evaluation, namely, the set of lotteries.

Each lottery a gives to each individual i a prospect a_i: a probability function over well-being levels. We are interested in individuals' preferences over prospects. We assume that they are orderings – that is, complete and transitive relations. Moreover, we assume that when restricted to degenerate prospects, these orderings match (the ordering of) well-being levels: a degenerate prospect is preferred to another just in case it results in greater well-being.

We will be asking whether our evaluations of policies should change upon variations in people's preferences over prospects – that is, changes in their risk attitudes in well-being. To address this, we need to consider different configurations of possible preferences over prospects that individuals might have. We will think of these as lists, with n entries, of orderings of prospects, and call them risk profiles, or simply profiles. The set of all risk profiles of interest will be denoted by $\mathcal{R}$. With $R = (R_1, \ldots, R_n) \in \mathcal{R}$, the i^{th} entry R_i stands in for individual i's ordering of prospects under that particular configuration. Finally, we are interested in evaluating what I will call *risk-encompassing social welfare functionals* (after Sen's 2017 notion of a social welfare functional):

> A **risk-encompassing social welfare functional** is a function that returns a social ordering of lotteries for each risk profile of interest.[8]

This is the framework we shall work with, and the assumptions made so far are taken as given.

We can now state a version of Harsanyi's aggregation theorem. In the present setup, the "*ex ante* Pareto" principle, according to which we should defer to rational people's unanimous preference for a given policy over another, can be stated as follows.

Strong Pareto.

> (**Informal**) Consider any lotteries a, b and profile R. If every individual ranks their prospect with a at least as high as their prospect with b in R, then a is at least as good as b in R. If, in addition, some rank their prospect with a strictly higher than their prospect with b in R, then a is better than b in R.

[7] As clarified in the Appendix, we restrict our focus to all probability functions with finite support, i.e. which assign non-zero probability to finitely many outcomes.

[8] For similar objects of study, see Blackorby *et al.* (2005: Ch. 7), Blackorby *et al.* (2008), Chambers and Echenique (2012), Fleurbaey and Zuber (2017) and Cibinel (2025b).

https://doi.org/10.1017/S0266267126100716 Published online by Cambridge University Press

This principle encodes a certain kind of respect for rational people's evaluation of their prospects, in line with their risk attitudes in well-being. We make the extra rationality assumptions, to be justified later, that individual and social preference satisfy the axioms of expected utility theory. We accordingly assume that all the entries of any risk profile in $\mathcal{R}$ satisfy the expected utility axioms and that our risk-encompassing social welfare functional outputs only social orderings that also satisfy the expected utility axioms. Individual and social preference orderings therefore admit of vNM utility representations. Our assumption that people's preferences for degenerate prospects match the ordering of well-being levels can now be stated as follows: individual vNM utility is a strictly increasing function of well-being.

Harsanyi's (1955) social aggregation theorem now allows us to infer that for all $R \in \mathcal{R}$, and *any* expected utility representations $U_1, \ldots, U_n$ for individual preferences in R, there exist positive real numbers $\alpha_1, \ldots, \alpha_n$ such that for all $a \in A$

$$U(a) = \alpha_1 U_1(a_1) + \ldots + \alpha_n U_n(a_n),$$

with U an expected utility representation for social preference. Since for any individual i, the function $U_i^\star = \alpha_i U_i$ is just another vNM utility function for i, another way of stating the result is this: for all $R \in \mathcal{R}$, there exist expected utility representations $U, U_1^\star, \ldots, U_n^\star$ for social and individual preferences respectively, such that for all $a \in A$,

$$U(a) = U_1^\star(a_1) + \ldots + U_n^\star(a_n).$$

When a lottery a assigns probability 1 to some outcome x, and so is equivalent to that outcome, we abuse notation and more conveniently write:

$$U(x) = U_1^\star(w_1) + \ldots + U_n^\star(w_n)$$

with $w_1, \ldots, w_n$ the well-being levels that $1, \ldots, n$ get in x.[9] This formulation reveals the Risk-Prioritarian nature of the sum of vNM utilities in the present setting: when it comes to evaluating outcomes, each person's well-being is treated differently depending on their risk attitude with respect to well-being.[10]

As things stand, the transformations $U_1^\star, \ldots, U_n^\star$ performed on people's well-being are individual-specific and profile-dependent. Individual-specificity means that two individuals with the same risk-attitude in well-being in some profile R may nonetheless be treated differently; profile-dependency means that two individuals

[9]Versions of this formula appear in Blackorby *et al.* (1980: 26, 28; 2008: 157), Broome (1991: 211); compare Mongin and Pivato (2016: 723), Gustafsson *et al.* (2023). Structurally, this is an additively separable representation: the term *generalized utilitarianism* is sometimes used for the special case where the transformations of well-being are the same (Blackorby *et al.* 2002: 559), and *transformed utilitarianism* for the general case where they may vary (Blackorby *et al.* 2008: 159). None of the authors cited defend the Risk-Prioritarian interpretation of the formula as the correct way to rank outcomes, as is my goal here.

[10]It bears emphasizing that these are risk attitudes with respect to well-being specifically. Since we are modelling risk attitudes via vNM expected utility theory (see Pratt 1964; Arrow 1971), there is no sense to be given to an individual's "risk attitude" *tout court*, without reference to a specific quantity/good (in contrast to normative interpretations of rank-dependent models, which can allow for a notion of an individual's general risk attitude; see Buchak 2013). When I omit the qualifier that a given risk attitude is in well-being, below, this is only done to avoid cluttering. The Risk-Prioritarian reliance on expected utility is justified later in this section (see also the Appendix).

https://doi.org/10.1017/S0266267126100716 Published online by Cambridge University Press

	Cam	Dale	Elise
a	w	w	w^-
c	w^-	w^+	w^-
b	w	w	w^-
d	w^+	w^-	w^-

Figure 1. Illustration of Interprofile Anonymity.

with the same risk-attitude in well-being in two different profiles R, R' may nonetheless be treated differently. These two features make room for unacceptable differential treatment, based on an individual's sheer identity.

To show that there can be a rationale behind the way in which a Risk-Prioritarian prioritizes people differently depending on risk attitude in well-being, I will now derive the existence of individual-independent and profile-independent transformations; I will call a risk-encompassing social welfare functional that is Risk-Prioritarian with individual-independent and profile-independent transformations *impartial*. Impartial Risk-Prioritarianism is characterized via the introduction of a second, new axiom, which I will call Interprofile Anonymity.

I will first illustrate how the axiom works with an example. Suppose that there are three individuals, Cam, Dale, and Elise. We consider two risk profiles, R and R'. Under R, Cam is risk-averse, Dale risk-seeking and Elise risk-neutral. Under R', Cam and Dale swap risk attitudes: Cam is risk-seeking (to the same degree that Dale is in R), Dale is risk-averse (to the same degree that Cam is in R), and Elise is again risk-neutral. Consider now four lotteries a, b, c, d with sure outcomes, where $w^+ > w > w^-$ are numbers representing quantities of well-being (Figure 1).

Notice that Elise is unaffected in a comparison between the four lotteries; so, we can forget about the last column in the table. Suppose c is socially better than a under R. This means that the gain of giving w^+ to risk-seeking Dale more than makes up for the loss of giving w^- to risk-averse Cam under R. R' differs only in that we swap Cam and Dale's risk attitudes. It stands to reason that the gain of giving w^+ to risk-seeking Cam should now, under R', outweigh the loss of giving w^- to risk-averse Dale, which amounts to d being better than b under R'. To think otherwise would effectively mean biasing one's social judgements in favour of Dale over Cam. Equal treatment of Cam and Dale, more broadly, requires a to be at least as good as c in R if and only if b is at least as good as d in R'.

To generalize this thought, we need to introduce two notions. First, an individual is "concerned" with respect to certain options just in case they potentially stand to gain or lose from a choice between those options. In the case just considered, Cam and Dale are concerned, while Elise is unconcerned. More precisely,

> if individual i's prospect with lottery a differs from at least one of their prospects with lotteries b, c, d, then i is defined to be *concerned* with respect to lotteries a, b, c, d.

Second, a "permutation" σ of individuals is a one-to-one correspondence between individuals. If permutation σ assigns, for example, Dale to Cam, then we say that Dale is Cam's counterpart (relative to that permutation). We can finally introduce the novel axiom of

https://doi.org/10.1017/S0266267126100716 Published online by Cambridge University Press

Interprofile Anonymity.

(**Informal**) Consider any lotteries a, b, c, d and profiles R, R'. Suppose there is some permutation such that, for all individuals (i) the prospect they get with a equals the prospect their counterpart gets in b; (ii) the prospect they get with c equals the prospect their counterpart gets in d; and moreover, for all concerned individuals, (iii) the ordering they get with R equals the ordering their counterpart gets in R'. Then a is at least as good as c in profile R if and only if b is at least as good as d in profile R'.

It is easy to check that Interprofile Anonymity indeed delivers the correct verdict in our illustration above, capturing the attractive thought that the *identity* of who has a certain risk attitude should not be morally significant.

Adding Interprofile Anonymity to Strong Pareto allows us to derive an individual-independent and profile-independent way of making tradeoffs between people with different attitudes towards risk in well-being, as shown in the Appendix. For all rational orderings of prospects, $R_*, R_\dagger, \ldots$, let $\mathbf{U}$ be any assignment of vNM utility representations $\boldsymbol{U}_{R_*}, \boldsymbol{U}_{R_\dagger}, \ldots$, to those orderings respectively. There will then exist positive real numbers $\alpha_{R_*}, \alpha_{R_\dagger}, \ldots$ such that for all $R \in \mathcal{R}$

$$U_R(a) = \alpha_{R_1}\mathbf{U}_{R_1}(a_1) + \ldots. + \alpha_{R_n}\mathbf{U}_{R_n}(a_n),$$

with R_i the ordering of individual i in profile R, and U_R a vNM utility function of social preference in R. The key point here is that the weights $\alpha_{R_*}, \alpha_{R_\dagger}, \ldots$ are chosen so as to scale vNM utility functions *before* the latter are assigned to particular individuals. Hence individuals with the same risk attitude in well-being, within and across profiles, are treated exactly the same – i.e. individual-specificity and profile-dependency are eliminated. Since for any ordering R_*, the function $\mathbf{U}^\star_{R_*} = \alpha_{R_*}\mathbf{U}_{R_*}$ is just another vNM utility function of R_*, the result can be stated in the following alternative format: for all rational orderings of prospects, $R_*, R_\dagger, \ldots$, there exists an assignment $\mathbf{U}^\star$ of vNM utility representations to these orderings, $\mathbf{U}^\star_{R_*}, \mathbf{U}^\star_{R_\dagger}, \ldots$, respectively, such that for all $R \in \mathcal{R}$,

$$U_R(a) = \mathbf{U}^\star_{R_1}(a_1) + \ldots. + \mathbf{U}^\star_{R_n}(a_n),$$

with R_i the ordering of individual i in profile R, and U_R a vNM utility function of social preference in R. The upshot for the question of how distributions of well-being should be evaluated is clear. For all profiles $R \in R$ and outcomes $x, \overline{x}$, with $w_1, \ldots, w_n$ the well-being levels that $1, \ldots, n$ get in x and $\overline{w_1}, \ldots, \overline{w_n}$ the well-being levels that $1, \ldots, n$ get in $\overline{x}$, x is socially at least as good as $\overline{x}$ under profile R just in case

$$\mathbf{U}^\star_{R_1}(w_1) + \ldots + \mathbf{U}^\star_{R_n}(w_n) \geq \mathbf{U}^\star_{R_1}(\overline{w_1}) + \ldots + \mathbf{U}^\star_{R_n}(\overline{w_n}).$$

Well-being distributions are evaluated by a sum of transformed well-being levels, where the transformations are vNM utility functions that are chosen in an individual-independent and profile-independent way. For ease of reference, I will call the vNM utility functions picked out by assignment $\mathbf{U}^\star$, which allow us to express social preference over outcomes as a sum of *unweighted* transformations on

https://doi.org/10.1017/S0266267126100716 Published online by Cambridge University Press

well-being levels, "privileged".[11] I note in closing that, as a bonus, Interprofile Anonymity and Strong Pareto together allow us to dispense with *assuming* expected utility theory at the individual and social level. In the Appendix, I adopt weaker rationality assumptions consistent with the well-known Allais preferences, and derive the version of the sum of vNM utility approach presented here – Impartial Risk-Prioritarianism – all the same.[12]

A comparison with related results in the spirit of Harsanyi's may be helpful at this juncture. Mongin and d'Aspremont (1998: 430–31) address the problem of profile-dependency by reconstructing Harsanyi's aggregation theorem in Sen's (2017) "multi-profile" setting of social welfare functionals (following Mongin 1994). They then show that an axiomatization of utilitarianism can be obtained by imposing standard axioms of anonymity ("Axiom A": 414) and independence of irrelevant alternatives ("Axiom I": 409). However, as they recognize, their multi-profile setting does not address the Sen–Weymark critique, since there is no distinction in their setup between vNM utility and well-being. By contrast, Blackorby *et al.* (2008: 157), incorporate the Sen–Weymark critique and derive profile-*dependent* "transformed utilitarian" social welfare functions – with the transformations given by vNM utility functions, much like we did immediately after assuming Strong Pareto above. However, they do not impose any axiom to ensure that the same transformation is applied to people with the same ordering of prospects, nor to relate different profile-dependent formulas in a normatively compelling way, as was achieved here, with the introduction of Interprofile Anonymity.[13]

The part of my axiomatic development which derives expected utility at the social and individual level from weaker axioms (see Appendix) is in the tradition of Mongin and Pivato (2015), McCarthy *et al.* (2020), Nebel (2020), Bradley (2022) and Gustafsson *et al.* (2023); of these authors, my strategy follows most closely Nebel and Bradley, the main difference being that I operate within a lottery rather than states-of-nature framework, and show that Outcome Anonymity, which they appeal to, is not essential to their approach to deriving expected utility. The connection with Gustafsson *et al.* (2023) is also worth mentioning. They do not assume that individuals' *ex ante* preferences are orderings, and yet are able to derive the conclusion that social preference is expectational with respect to a sum of transformed well-beings, using weak dominance axioms. Their only individual-level assumption in the fixed-population setting, which is the one used in this paper, is stochastic dominance.[14] In comparison, I do here assume that individuals rank prospects with full-blown orderings; this allows me to operate in a multi-profile

[11]The uniqueness of the "privileged" assignment $\mathbf{U}^{\star}$ is as follows. For $\alpha > 0$ and $\beta, \beta', \ldots \in \mathbb{R}$, all and only assignments $\mathbf{U}^{\star\star}$ of the following form would also count as privileged: R_* gets associated to $\mathbf{U}^{\star\star}_{R_*} = \alpha \mathbf{U}^{\star}_{R_*} + \beta$, $R_\dagger$ gets associated to $\mathbf{U}^{\star\star}_{R\dagger} = \alpha \mathbf{U}^{\star}_{R\dagger} + \beta'$, ..., with α constant. See Appendix.

[12]I leave for future research a line of inquiry pointed out by a referee, namely exploring the possibility of weakening Strong Pareto in the present setup. For extant results, see Adler (2019: Ch. 4), (2025: Ch. 7) and Gustafsson *et al.* (2023).

[13]Their (unrestricted) multi-profile result is instead an impossibility theorem (2008: 161), as are those of Blackorby *et al.* (2002: 587; 2005: 230). The Risk-Priority View escapes these impossibilities by denying welfarism (in its profile-independent form; see Cibinel 2025b).

[14]In their variable population result, they adopt a weakening of individual stochastic dominance.

https://doi.org/10.1017/S0266267126100716 Published online by Cambridge University Press

setting, and explore how the individual transformations on well-being relate within and across profiles, *depending on* what orderings individuals adopt. While Gustafsson *et al.* (2023) effectively leave individual preferences out of their framework (an apt approach given their aim), such preferences are of central interest to the present project. As a consequence, the formula they derive is less committal than the one we obtained above. For example, their result is consistent with Utilitarianism and Prioritarianism, while mine rules such views out as general rankings of well-being distributions.

3. Scaling of vNM Functions

The crucial question now to be considered is how exactly vNM utility functions should be scaled in the framework of section 2. This question arises for *all* proponents of the view that social value is given by a sum of vNM utilities. But it means something different on a Risk-Prioritarian interpretation of such a sum than it does on a utilitarian interpretation. On a utilitarian interpretation of Harsanyi's aggregation theorem, scaling vNM utility functions amounts to deciding how to make comparisons of well-being improvements and losses for different people (see Adler 2019: 55–64; Fleurbaey and Zuber 2021). On the Risk-Prioritarian interpretation, scaling vNM utility functions amounts to deciding how to make comparisons of the moral significance of well-being improvements and losses for people with different risk attitudes in well-being.[15]

What the Risk-Prioritarian interpretation of Harsanyi's aggregation theorem developed in section 2 gives us is thus a *class* of theories, which differ from one another in the way they make interpersonal comparisons of vNM utility for people with different risk attitudes in well-being. There is an analogy here with Prioritarianism. The Priority View is also a class of theories, with each member of this class differing from the others in the degree to which it prioritizes worse-off individuals. Some prioritarian theories implausibly give too much – or too little – weight to the worse off. While the class of Prioritarian rankings can be characterized axiomatically (Adler 2022: 63), particular normative judgements have to guide prioritarians in making their choice of which version of their theory to accept.

[15]Of course, utilitarians see no difference between quantities of well-being and how much these quantities count (Broome 1991). But non-utilitarians disagree: according to a prioritarian, for example, an equal improvement in well-being matters more, morally, the worse off an individual is (Parfit 2002; Adler 2011). Greaves (2015: 587) calls "modern utilitarianism" the view that vNM utility equals (what I have been calling) moral significance. On this taxonomy, both Harsanyi's Utilitarianism and the Risk-Priority View count as versions of "modern utilitarianism". It seems to me that most modern utilitarians in this sense are either squarely followers of Harsanyi, or at any rate fail to take seriously the idea that there could be quantities of well-being independently of vNM utility (see Greaves 2017). A case that I find hard to categorize is Hammond (1996). His theory sums individual vNM utilities, where these are derived from ethical choices that bear upon each individual singly; unlike Harsanyi, he explains that these quantities are not what classical utilitarians have in mind, and rather have "purely ethical significance" (94). He nonetheless calls each individual's vNM utility so understood their "welfare" (94). I am unsure how many "modern utilitarians" would side with the Risk-Priority View *if* they became convinced of its assumption that we can make sense of quantities of well-being independently of vNM utility, and how many would instead cling to the truly utilitarian aspect of their position and side with Utilitarianism. As I hope to clarify with this paper, the two alternatives have very different motivations.

https://doi.org/10.1017/S0266267126100716 Published online by Cambridge University Press

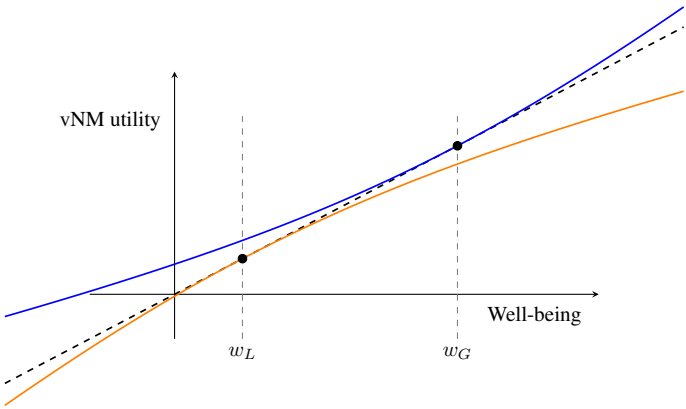


Figure 2. Illustration of vNM utility scaling as a function of well-being, as discussed in main text.

With these two analogies in mind – utilitarian vNM scaling and prioritarian choice of transformation function – I now present a particular approach for a Risk-Prioritarian scaling of vNM utility functions.[16] The aim here is to put forward a relatively concrete proposal as proof of concept in the service of an argument for the viability of Risk-Prioritarianism in general. It is not to conclusively advocate for this particular solution.

We identify two reference levels of well-being (according to our well-being function): one, w_L, corresponding to the state of someone who is moderately badly off, and another, w_G, corresponding to the state of someone who is moderately well off. To be clear, this selection of reference levels is based on the actual condition of each individual, i, in outcomes which the welfare function maps to w_L and w_G, respectively, for i; w_L and w_G are just *numbers* returned by the welfare function W, and have no independent moral significance. The Risk-Prioritarian scaling to be laid out prioritizes (i) individuals who are risk-averse in well-being over individuals who are not risk-averse in well-being below level w_L (circumstances are "dire"); (ii) individuals who are risk-neutral in well-being over risk-aversion and risk-seekingness in well-being between w_G and w_L (circumstances are "normal"); (iii) individuals who are risk-seeking in well-being over risk-aversion and risk neutrality above w_L (circumstances are "favourable").

Here is how the scaling in question selects privileged vNM utility functions (Figure 2). We start by arbitrarily selecting a vNM function representing the ordering of prospects R_N that is risk-neutral in well-being, $\mathbf{U}^*_{R_N}(w) = \alpha w + \beta$. For any risk-averse ordering of prospects R_A, the privileged vNM function $\mathbf{U}^*_{R_A}$ is selected so that its first derivative at w_L equals α, i.e. $\mathbf{U}^{*\prime}_{R_A}(w_L) = \alpha$. For any risk-seeking ordering of prospects R_S, the privileged vNM function $\mathbf{U}^*_{R_S}$ is selected so that its first derivative at w_G equals α, i.e. $\mathbf{U}^{*\prime}_{R_S}(w_G) = \alpha$.

[16]My proposal essentially reinterprets Fleurbaey and Zuber's (2021) utilitarian scaling in a Risk-Prioritarian key. As is common in the literature on utilitarian scaling of vNM utility, I focus on vNM functions that are constantly risk-averse, risk-neutral or risk-seeking, and differentiable; see again Fleurbaey and Zuber (2021).

https://doi.org/10.1017/S0266267126100716 Published online by Cambridge University Press

An intuitive case for this scaling can be made along the following lines. There are different ways to pursue one's good (well-being) under risk: risk-averse individuals focus on avoiding worse case scenarios; risk-seeking individuals focus on pursuing better case scenarios. For this reason, it is somewhat worse if risk-averse people end up very poorly off, compared to risk-seeking ones. And it is to some extent better if risk-seeking people end up very well off, compared to risk-averse ones. Crucially, it is not worse *for* the individuals in question: on the Risk-Prioritarian (unlike on the utilitarian) interpretation of vNM utility, cardinally measurable and comparable well-being levels are what people's vNM utility functions *take as arguments*. Rather, the claim is that these outcomes are worse or better from a moral or social point of view (compare Nebel 2022: 39). Thus, for instance, going from w_L to $w_L - \Delta$ is exactly as bad for a risk-averse individual as it is for a risk-seeking individual, by design of our section 2 setup. The Risk-Prioritarian scaling here considered, however, sees a risk-averse individual going from w_L to $w_L - \Delta$ as morally worse than a risk-seeking individual going from w_L to $w_L - \Delta$.[17]

4. The Risk-Priority *Ethos*

While Interprofile Anonymity sweetens the pill of Risk-Prioritarian violations of Outcome Anonymity and the Suppes–Sen Principle, the task remains of providing a philosophically compelling story for why people's well-being should be weighed differently, according to their risk attitude, *even in risk-free cases*. Fundamentally, what motivates taking account of risk attitudes when there is no risk involved? Ostensibly, Strong Pareto only recommends giving weight to risk attitudes in risky contexts. The counterintuitive Risk-Prioritarian violations of Outcome Anonymity and the Suppes–Sen Principle, by contrast, result from giving weight to risk attitudes in risk-*free* contexts. What philosophical move could possibly get us from the former stance to the latter?

This section pursues two interrelated answers, concerning justifiability to each person and individual responsibility, respectively. To keep things simple, I will focus on violations of Outcome Anonymity; my remarks generalize to violations of the Suppes–Sen Principle. The strategy is this: I will introduce cases that proponents of Strong Pareto should already recognize as supporting some form of differential weighting of well-being, and describe how these verdicts are typically motivated. I will then show that such standard justifications of these verdicts should be developed and built upon in ways that lead to more radical violations of Outcome Anonymity, involving risk-free contexts.

It will be helpful to be able to refer to a couple of risk-free cases in what follows. Hence, consider

[17]Fleurbaey and Zuber (2021) follow the Harsanyi paradigm, whereby there is no antecedent cardinal measure of well-being, which is instead taken to be provided by vNM utility (see Fleurbaey and Zuber 2022: 463). However, at one point they claim that "[w]e could interpret income as a comprehensive and interpersonally comparable measure of *ex post* well-being" (2021: 378). This comment and the brief ensuing discussion (378–9) foreshadow a version of the approach I adopt here. See also their (Fleurbaey and Zuber 2017) for an approach more in line with mine.

https://doi.org/10.1017/S0266267126100716 Published online by Cambridge University Press

Dire Circumstances, in which Alda is risk-seeking and Bruno risk-averse. They are both poorly off, at well-being level $w_L - \Delta$. A decision-maker can either (A) improve Alda's well-being to w_L, or (B) improve Bruno's well-being to w_L.

And compare

Favourable Circumstances, in which risk-seeking Alda and risk-averse Bruno are both well-off, living happily at well-being level w_G. A decision-maker can either (C) improve Alda's well-being to $w_G + \Delta$, or (D) improve Bruno's well-being to $w_G + \Delta$.

These two cases are stipulated to be as risk-free as possible. In each case, Alda and Bruno are strangers to the decision-maker; do not know that one of them is going to be benefited; and their current well-being level is *not* the result of some previous decision to take or turn down a gamble. The decision-maker's choice of who to give the well-being gain to is entirely based on which ranking of well-being distributions would be better impartially from a social or moral point of view. The Risk-Prioritarian sum of vNM utilities developed in section 2, as scaled in section 3, nonetheless recommends (counterintuitively) the non-anonymous verdicts that (B) should be chosen over (A), and (C) over (D).

4.1 Justifiability

To begin with our first justification of these Risk-Prioritarian verdicts, contrast Dire and Favourable Circumstances with a pair of cases in which risk plays an obvious role. In

Endorsed Risk, risk-seeking Alda ends up poorly off at well-being level $w_L - \Delta$, as a result of a risky policy that she wished the decision-maker would implement. Risk-averse Bruno is at well-being level w_L.

In a second case

Unendorsed Risk, risk-averse Bruno ends up poorly off at well-being level $w_L - \Delta$, as a result of a risky policy that he wished the decision-maker would not implement. Risk-seeking Alda is at w_L.

Intuitively, Alda's predicament in Endorsed Risk is less regrettable, morally, than Bruno's predicament in Unendorsed Risk. A natural explanation of this judgement appeals to the idea of justifiability to each individual affected (Scanlon 1998). In Endorsed Risk, Alda cannot reasonably blame the decision-maker for her unfortunate condition, since the decision-maker's choice perfectly reflected Alda's own evaluation of her prospects. When Alda ends up unlucky, the decision-maker can justify the decision to Alda on the basis of her preference *ex ante* (James 2012; Frick 2015). By contrast, Bruno appears to have grounds for complaint in Unendorsed Risk, since his unfortunate condition is due to the decision-maker enacting a policy that went against his own evaluation of his prospects.

https://doi.org/10.1017/S0266267126100716 Published online by Cambridge University Press

These cases show how a commitment to respecting people's preferences *ex ante* in the way that Strong Pareto prescribes already entails some differential evaluation, resulting in violations of Outcome Anonymity (Cibinel 2022). In both Endorsed Risk and Unendorsed Risk, there is one individual at well-being level $w_L - \Delta$ and one at w_L: yet, the outcome of Endorsed Risk is better than that of Unendorsed Risk.

It would be a further step, however, to argue for differential treatment in a context of zero risk: if we can benefit only one of Alda or Bruno under conditions of certainty, proponents of *ex ante* Pareto would typically say that we should be indifferent between the two options.[18] If the decision-maker chooses (B) in Dire Circumstances, Alda is not badly off as a result of a gamble that she was willing to take, but rather because of a decision the decision-maker took with full information of its consequences. Surely, one might think, in such a risk-free situation (A) and (B) are to be deemed equally good morally speaking, in line with Outcome Anonymity.

I will now argue that this more radical violation of Outcome Anonymity can nonetheless be vindicated by considerations of justifiability to everyone involved. In other words, there is a viable route that takes us from the relatively uncontroversial

> ***Ex Ante* Intuition (I)**, according to which Alda ending up poorly off in Endorsed Risk is not as bad as Bruno ending up poorly off in Unendorsed Risk,

all the way to the verdicts of the Risk-Priority View in risk-free cases also: for example, the verdict that the decision-maker should help Bruno rather than Alda in Dire Circumstances.

Notice, first, that whether an act is justifiable to every individual affected depends on the cumulative effects that can be expected, overall, if similar acts are performed in similar circumstances (Scanlon 1998: 203–4). That is, the justifiability status of specific acts hinges on the justifiability status of principles "for the general regulation of behavior" (Scanlon 1998: 153) – in particular, for the regulation of behaviour in related conditions. If a policy turns out to someone's disadvantage on some specific occasion, but is shown to be part of a broader scheme that can be expected to further this person's interest, it seems that the person has no reasonable ground for rejecting the specific act. In light of this clarification, we can ask: how would a general scheme of treating everyone equally in all circumstances, as Outcome Anonymity demands, fare against a Risk-Prioritarian plan of differential treatment?

To answer this question we need to look at what would happen if Alda, Bruno and the decision-maker *were* to settle on a general principle for how to proceed should they find themselves in Dire Circumstances or in Favourable Circumstances, with no evidence making one scenario more likely than the other to obtain. On an Outcome Anonymity scheme, the decision-maker should be indifferent between whom to help in either circumstance. They would thus follow the plan:

> (**S1**) Be indifferent between (A) and (B) if Dire Circumstances obtains, and be indifferent between (C) and (D) if Favourable Circumstances obtains.

[18]See e.g. James (2012: 279), who would recommend flipping a fair coin, following Diamond (1967); I will momentarily operationalize indifference in a similar way.

https://doi.org/10.1017/S0266267126100716 Published online by Cambridge University Press

On a Risk-Prioritarian scheme, the decision-maker should prioritize risk-averse Bruno if circumstances are dire, and risk-seeking Alda if circumstances are favourable. That is, the decision-maker's scheme should be:

(**S2**) choose (B) if Dire Circumstances obtains, and choose (C) if Favourable Circumstances obtains.

Which scheme would the group agree to enact?

Consider first S1. If Dire Circumstances obtains, the decision-maker is indifferent as to who should be benefited. For Alda and Bruno, this amounts to equal (and perfectly inversely correlated) chances of getting well-being level w_L or well-being level $w_L - \Delta$. (Anything other than equal chances seems to encode bias in favour of one individual, and is thus incompatible with indifference.) If Favourable Circumstances obtains, the decision-maker will again decide who should be benefited on the basis of indifference. This gives Alda and Bruno equal (and perfectly inversely correlated) chances of getting well-being level w_G or well-being level $w_G + \Delta$. So, S1 looks exactly the same from Alda's perspective as it does from Bruno's. By contrast, S2 is more risky for Alda and more safe for Bruno. S2 doubles Alda's chance at getting well-being $w_G + \Delta$, since if the decision-maker is presented with Favourable Circumstances, they will benefit Alda for sure. However, S2 also doubles Alda's chance at well-being $w_L - \Delta$, since if the decision-maker is presented with Dire Circumstances, they will benefit Bruno for sure. Alda's chances of obtaining well-being w_L or w_G are correspondingly reduced to zero. By contrast, S2 doubles Bruno's chances at well-being levels w_L and w_G, and reduces to zero his chance to obtain either well-being level $w_L - \Delta$ or $w_G + \Delta$.

Alda, being risk-seeking, will prefer S2 to S1: in expectation, the two schemes give her equal future well-being, but the former increases her chance at the very best outcome. Bruno, being risk-averse, will also prefer that the decision-maker adopt S2 instead of S1: in expectation, the two schemes give him equal future well-being, but the former reduces his chance at the worst outcome. Since S2 would be unanimously preferred, it seems that the decision-maker can justify its adoption to all, while S1 can be reasonably rejected. In particular, the decision-maker can justify choosing (B) to Alda if Dire Circumstances in fact obtains, since it is part of a *general* scheme that she has no reason to object to.[19] This example illustrates how, by construing risk-free choices as governed by general principles for how to choose in a *range* of interrelated scenarios, Risk-Prioritarian verdicts can be justified to every individual affected, along similar lines as was done for *Ex Ante* Intuition (I). In the long run, implementing a Risk-Prioritarian scheme instead of one conforming to Outcome Anonymity better satisfies people's preferences for risks in well-being: it gives more gambles (in well-being) to more risk-seeking individuals, and more safety (with respect to well-being) to more risk-averse individuals. As such, it seems no one could object to such a scheme.[20]

[19]Thomson (1990: 195) uses an argument with a similar structure to conclude that some versions of the trolley case are "exceptions to the general rule that one may not kill one to save five".

[20]It bears emphasizing that the reason why everyone would rationally accept such a scheme is that it better lines up with their preferences over well-being lotteries, and *not* that it is good for them, in the sense of improving everyone's actual well-being – for some relevant discussion, see Cibinel (2025b: 532).

https://doi.org/10.1017/S0266267126100716 Published online by Cambridge University Press

A ranking of policies in terms of their goodness from a social or moral point of view is simply a general principle specifying how a decision-maker should act in certain given circumstances they might find themselves in. This suggests that a Risk-Prioritarian ranking can be justified to the individuals involved, whereas a ranking that respects Outcome Anonymity cannot.

4.2 Responsibility

Justifiability to each person is one of two main ways in which Strong Pareto is typically motivated. The other, related approach appeals to individual responsibility.[21] Consider

> *Option & Brute Luck*, where both Risk-seeking Alda and risk-averse Bruno end up poorly off at well-being level $w_L - \Delta$. Alda's bad outcome, however, is the result of a gamble that she chose of her own will, whereas Bruno's bad outcome is the result of a gamble that he had no choice but to accept.

It seems right to say that Alda's predicament is less regrettable, from a moral point of view, than Bruno's predicament. Both events are bad, but Bruno's bad luck is worse. A natural explanation involves the thought that individuals should be held accountable for risks that they freely choose to incur (Dworkin 1981). When Alda ends up poorly off as a result of a risk that she decided to take, the thought goes, she is in some sense responsible for her unfortunate condition – this is a case of what Dworkin calls "option luck".[22] The same cannot be said of Bruno, who ends up badly off through no fault of his own – this is a case of what Dworkin calls "brute luck". Suppose we can either improve Alda's well-being level to w_L, or else Bruno's well-being level to w_L. Since Bruno's predicament is more regrettable, we should help Bruno instead of Alda, violating Outcome Anonymity.

When such a differential treatment is defended, however, it is not extended to cover the more radical violations of Outcome Anonymity entertained by the Risk-Priority View, involving risk-free cases. In other words, there is no suggestion in this literature that we should prioritize Bruno over Alda in a case like Dire Circumstances, where neither Alda nor Bruno are in any obvious sense responsible for the unfortunate condition they find themselves in.[23] I will now argue that differential prioritization in risk-free cases is nonetheless justified by considerations broadly

[21]As noted by Fleurbaey (2010: 675), remarks by Hammond (1983) and Kolm (1998) suggest such a picture connecting respect for *ex ante* preferences to personal responsibility for bearing the consequences of risks incurred; I will discuss Dworkin (1981).

[22]The flipside of this is that if she turns out well off instead, she should be entitled to the outcome of her good luck. In this way, as Dworkin (1981) puts it, we can say to the risk-seeker that "the possibility of loss was part of the life they chose – that it was the fair price of the possibility of gain" (294). This passage is also cited by Fried (2003: 133), who offers an extensive critical discussion of Dworkin's position.

[23]Brute-luck egalitarians precisely want to neutralize the effects of such "brute luck" (see e.g. Rakowski 1991: 74 – leaving aside the complication that Rakowski, like Dworkin and some other luck egalitarians, is interested in equality of resources and not of well-being).

https://doi.org/10.1017/S0266267126100716 Published online by Cambridge University Press

based on individual responsibility. In other words, there is a viable route that takes us from the

> ***Ex Ante* Intuition (II)**, according to which, in Option & Brute Luck, the fact that Alda ends up poorly off is not as bad as the fact that Bruno ends up poorly off,

all the way to the more controversial verdicts of the Risk-Priority View.

To begin, it will help to think about how background risks and opportunities shape our judgements about responsibility for risks incurred. Suppose that, in

> *Background Risk*, there is a 50% chance that both Alda and Bruno will end up poorly off at well-being $w_L - \Delta$. Such a fact is given; there is nothing they can do about it. However, they can affect what happens if that bad outcome does not obtain. They can either have well-being level w_L in all remaining possibilities, or else well-being level $w_L - \Delta$ and well-being level $w_G + \Delta$ with equal probability. Risk-seeking Alda chooses the latter option, and risk-averse Bruno the former. As it happens, Alda ends up poorly off at well-being level $w_L - \Delta$.[24]

Is Alda here responsible for her unfortunate condition, in the way she is in Option & Brute Luck? I think we should feel ambivalent about how to answer this question, and that this supports a comparative understanding of individual responsibility for risks, which in turn counts in favour of the Risk-Priority View. Let's unpack this line of argument.

It might be tempting to say that whether Alda is responsible in Background Risk depends on whether she is at well-being level $w_L - \Delta$ because of the 50% background risk, or rather because of the added 25% risk she freely chose to incur. But this proposal does not stand up to scrutiny. If the mechanism that caused the background risk is exactly the same one that also caused the extra risk, it is not clear what it would mean for Alda to have ended up poorly off because of one or the other "portion" of the total risk she faced. It might be suggested that we need to know what would have happened had she chosen otherwise: if, had she chosen in the same way as Bruno, she would have ended up safe at well-being level w_L, then she is responsible for her actual bad condition; but if, had she chosen in that way, she would still have ended up at well-being level $w_L - \Delta$, then she is not responsible for her actual bad condition.

But, first, there may be no fact of the matter as to what would have happened had Alda chosen otherwise (see Hare 2011, 2015). And, second, answering this question does not seem normatively relevant after all. Suppose the 50% background chance that Alda and Bruno end up at well-being level $w_L - \Delta$ in Background Risk is due to the presence of an incoming storm, which carries the threat of being hit by lightning.

[24]That background risks complicate Dworkin's distinction between option and brute luck is noted by Knight (2021) and Lippert-Rasmussen (2023). Neither of these authors, however, suggests that considerations of background risk point in the direction of the Risk-Priority View, which will be the main claim of this subsection.

https://doi.org/10.1017/S0266267126100716 Published online by Cambridge University Press

Each of them can stay at home: lightning will hit their house with 50% probability, but there is nowhere more protected where they can go. If their house remains intact, they will be safe but bored, at well-being level w_L. Alternatively, they can go out surfing. If all goes well, they enjoy themselves, at well-being level $w_G + \Delta$; however, going surfing increases the chance of being hit by lightning to 75%. Alda goes surfing and gets hit by lighning. Suppose that Bruno, who stays home, is also hit by lightning. If we only have the resources to provide first-aid care to one of them, who should we help? I think the answer is that Bruno should be prioritized. And this is so even if we learn *ex post* that, had Alda remained at home, she would have been hit by lightning anyway. What matters to assessing Alda's situation, it seems to me, is that she chose to run more risk than Bruno did.[25]

The discussion does show, nonetheless, that the sense in which Alda is responsible for her bad condition in Background Risk is rather *sui generis*, and quite distinct from the more familiar notion of moral responsibility employed in attributions of praise and blame. Alda's odds were not ideal to start with, since she faced a 50% background risk that she could not get rid of. And it's not as if she wanted risk for its own sake: had she had the option of securing well-being level $w_G + \Delta$ with certainty – going surfing with a clear sky – she would have chosen that. Taking a gamble means that we are acting in sub-optimal circumstances, which must be taken into account when evaluating responsibility for risks incurred. This points to the suggestion that we interpret Alda's responsibility for her bad condition comparatively, to mean that she should receive less priority, in this case, than people in similar circumstances who took fewer risks, and that the reason for this is precisely that she chose to be more risk-seeking than those other people. That choice was hers, and in this sense she is responsible for it. But this does not mean that, if we can help her at no cost to others, we have no reason to do so (cf. Lippert-Rasmussen 2023: sect. 7).

We can now notice, however, that this consideration also applies to Alda's situation in Option & Brute Luck: the reason why Alda took a gamble in that case must have been that her options were somewhat limited. If she could have acted so as to achieve the best possible outcome of that gamble with certainty, she would have done that. So, a comparative understanding of responsibility for risks incurred seems to be appropriate quite generally.

We next need to combine what we learned from Background Risk with the following observation: the choice of what risk attitude in well-being, or vNM utility function, to adopt is itself a *risky* choice made by the individual. To elaborate, let's begin by noting that preferences are dispositions to choose. Preferences under risk are therefore dispositions to choose when we don't know the consequences of our actions. They can thus be understood as complete plans for how to act in any given circumstance. Once I have chosen such a plan and face some given decision problem, all I have to do is implement an algorithm: pick an option that maximizes my vNM utility.[26] In making the choice of risk attitude – that is, in ranking all

[25]An analogy may be noted here with so-called "Frankfurt cases" in the literature on moral responsibility (Frankfurt 1969); however, I am about to sketch some differences between moral responsibility and the kind of responsibility for risks incurred of interest here.

[26]Surely, it can be argued, this is at best a simplification: people do not choose their risk attitude in well-being once and then stick to it throughout their lives. I grant the point. A more elaborate treatment would

https://doi.org/10.1017/S0266267126100716 Published online by Cambridge University Press

possible lotteries over well-being levels – each individual will take account of background risks and opportunities they might encounter: perhaps I will trip and break my neck; perhaps I will win the lottery; and so on. Since we know about these risks and opportunities when we choose our risk attitude, they are relevantly like the 50% background risk that Alda faces in Background Risk. If she is responsible for her condition then – which she is, if we want to hold her responsible in Option & Brute Luck – so too are those of us who choose to be risk-seeking and end up poorly off "through no fault of their own".

Ultimately, then, my discussion should be interpreted as casting doubt on the usefulness of Dworkin's (1981) distinction between brute and option luck.[27] Take a paradigm case of brute luck: a rock falls from the sky through the roof and hurts a given individual, Frank, while he was resting in his house, leaving him at $w_L - \Delta$ when he would otherwise have had well-being level w_L. Against my proposal of prioritizing Frank's predicament differently, depending on his risk attitude in well-being, one might note: "there is nothing Frank could have done to alter the probability of this happening!" In response, I claim: the normatively significant event is not that a rock hits Frank's head, but rather that Frank ends up at well-being $w_L - \Delta$; and there is something that Frank could have done to alter the probability of the claim that *Frank ends up at well-being level* $w_L - \Delta$. Before the misfortune, we should leave open that Frank's future might present him with a risky choice involving the possibility of a Δ-loss from w_L. If, prior to the rock falling, we assume that Frank is risk-seeking in well-being, that will result in a *different* (higher) risk of ending up at $w_L - \Delta$ in the future than if we assume that Frank is risk-averse in well-being. This is my basis for claiming that it is comparatively more urgent to help Frank if he is risk-averse in well-being in this case.

In sum, background risks are part of the context within which we make any choice, including our choice of general policy for how risky to be. Recognition of this fact changes our assessment of any misfortune that befalls a given individual in two opposing ways. On the one hand, we will be able to identify some point in time relative to which the claim that "a misfortune of this magnitude eventuates" was a background risk, and also able to identify some choice the individual could have made at that time – including choice of risk attitude in well-being – which would have altered the probability of that risk. On the other hand, the fact that this claim had positive probability to start with is obviously beyond the individual's control, and this affects the sense in which they can be held responsible for it. The brute/

take account of the possibility of changes in vNM utility function (i.e. preferences under risk) for an individual. For that matter, a complete treatment should also speak to the possibility that an individual may not be modelled as having a vNM utility function at all. These two restrictive assumptions, however, are common in the literature based on Harsanyi, where vNM utility is taken to measure well-being (see Adler 2019: 55–64, for a summary). As noted in the Introduction, my aim here is not to provide a comprehensive defence of the sum of vNM utility approach, but to suggest how it might incorporate the Sen–Weymark critique systematically and be defended vis-à-vis the resulting counterintuitive verdicts. For present purposes, then, I acknowledge the limitation here highlighted and leave a fuller treatment of the issue for future research. Thanks to a referee for pressing me on this issue.

[27]For a similar conclusion, see Bognar (2019); note that Dworkin himself points out that option and brute luck come in degree, and that the latter can be turned into the former through insurance (1981: 293). However, he hastens to add that this "does not erase the distinction" (293).

https://doi.org/10.1017/S0266267126100716 Published online by Cambridge University Press

option luck distinction attempts to demarcate a clear line between personal responsibility for one's condition and lack thereof; but there isn't one. The erosion of this distinction, I suggest, points to a revision to the notion of personal responsibility for one's condition: it is to be cashed out comparatively, in terms of what degree of priority one is entitled to receive in the given circumstances, relative to other individuals.

Thus, Risk-Prioritarian verdicts such as those in Dire and Favourable Circumstances can be supported by considerations of personal responsibility similar to those that motivate the *Ex Ante* Intuition (II). A Risk-Prioritarian ranking of policies in terms of how good they are socially or morally speaking systematizes such judgements of prioritization based on individual responsibility.

5. Conclusion

While weighing people's well-being by risk attitudes may seem odd at first, I have tried to convey a way of thinking about distributive ethics which makes the proposal quite natural. This approach is impartial in just the right way: enough as to not play favourites, but not so much that it disregards people's preferences. It can be motivated by interrelated considerations of justifiability to each person and individual responsibility. And it illuminates the philosophical significance of Harsanyi's social aggregation theorem, while paying heed to the Sen–Weymark critique. I conclude that the Risk-Priority View offers an attractive account for how well-being should be aggregated.

Acknowledgements. I would first like to thank Lara Buchak and Jake Nebel, both of whom helped me massively at all stages of development of the project. I thank Chris Bottomley, Richard Bradley, Franz Dietrich, Adam Elga, Daniel Herrmann, Kacper Kowalczyk, Marcus Pivato, Jeff Russell, Weng Kin San, Margaret Shea, Lingzhi Shi, Teruji Thomas, John Weymark and Qichen Yan for beneficial discussion, and Harvey Lederman and Sarah McGrath for that as well as for very useful written comments. For constructive feedback, my gratitude also goes to a number of anonymous referees (in particular two referees of *Economics and Philosophy*) and to audiences at the 2024 Global Priorities Workshop and 2024 Meeting of the Society for Social Choice and Welfare, where a talk version of this paper was presented.

References

Adler M.D. 2011. *Well-Being and Fair Distribution: Beyond Cost-Benefit Analysis*. Oxford: Oxford University Press.

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

Adler M.D. 2022. Theory of prioritarianism. In *Prioritarianism in Practice*, ed. M.D. Adler and O.F. Norheim, 37–127. Cambridge: Cambridge University Press.

Adler M.D. 2025. *Risk, Death, and Well-Being: The Ethical Foundations of Fatality Risk Regulation*. Oxford: Oxford University Press.

Adler M.D., M. Ferranna, J.K. Hammitt and N. Treich 2021. Fair innings? The utilitarian and prioritarian value of risk reduction over a whole lifetime. *Journal of Health Economics* **75**, 102412.

Adler M.D. and N. Holtug 2019. Prioritarianism: a response to critics. *Politics, Philosophy and Economics* **18**, 101–144.

Allais M. 1953. Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'ecole americaine. *Econometrica* **21**, 503–546.

Arrow K.J. 1971. *Essays in the Theory of Risk-bearing*. Chicago: Markham Publishing Co.

https://doi.org/10.1017/S0266267126100716 Published online by Cambridge University Press

Bell D.E. and H. Raiffa 1988. Marginal value and intrinsic risk aversion. In *Decision Making: Descriptive, Normative, and Prescriptive Interactions*, ed. D.E. Bell, H. Raiffa and A. Tversky, 384–397. Cambridge: Cambridge University Press.

Blackorby C., W. Bossert and D. Donaldson 2002. Utilitarianism and the theory of justice. In *Handbook of Social Choice and Welfare: Volume* **1**, ed. K.J. Arrow, A.K. Sen and K. Suzumura, 543–596. Dordrecht: Elsevier.

Blackorby C., W. Bossert and D. Donaldson 2005. *Population Issues in Social Choice Theory, Welfare Economics, and Ethics*. Cambridge: Cambridge University Press.

Blackorby C., D. Donaldson and J.A. Weymark 1980. On John Harsanyi's defences of utilitarianism. *Discussion Paper of the Center for Operations Research and Econometrics, Université Catholique de Louvain 8013*, 1–40. https://john-weymark.github.io/files/BDW80CORE.pdf.

Blackorby C., D. Donaldson and J.A. Weymark 2008. Social aggregation and the expected utility hypothesis. In *Justice, Political Liberalism, and Utilitarianism: Themes from Harsanyi and Rawls*, ed. M. Fleurbaey, M. Salles and J.A. Weymark, 136–183. Cambridge: Cambridge University Press.

Bognar G. 2019. Catering for responsibility: brute luck, option luck, and the neutrality objection to luck egalitarianism. *Economics and Philosophy* **35**, 259–281.

Bradley R. 2022. Impartial evaluation under ambiguity. *Ethics* **132**, 541–569.

Broome J. 1991. *Weighing Goods: Equality, Uncertainty and Time*. Oxford: Blackwell.

Broome J. 2004. *Weighing Lives*. Oxford: Oxford University Press.

Broome J. 2015. General and personal good: Harsanyi's contribution to the theory of value. In *The Oxford Handbook of Value Theory*, ed. I. Hirose and J. Olson, 249–66. Oxford: Oxford University Press.

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.

Cibinel P. 2022. Risk attitudes and justifiability to each. *Ethics* **133**, 106–121.

Cibinel P. 2025a. *Individual Risks and their Social Outcomes*. Ph.D. thesis, Princeton University.

Cibinel P. 2025b. Welfare and autonomy under risk. *Philosophy and Phenomenological Research* **110**, 526–551.

Diamond P.A. 1967. Cardinal welfare, individualistic ethics, and interpersonal comparison of utility: Comment. *Journal of Political Economy* **75**, 765–766.

Dietrich F. Forthcoming. Welfare vs. utility. *Economic Theory*. https://philpapers.org/rec/DIEWVU.

Dworkin R. 1981. What is equality? Part 2: Equality of resources. *Philosophy and Public Affairs* **10**, 283–345.

Epstein L.G. and U. Segal 1992. Quadratic social welfare functions. *Journal of Political Economy* **100**, 691–712.

Ferranna M., J.K. Hammitt and M.D. Adler 2023. Age and the value of life. In *The Routledge Handbook of the Economics of Ageing*, ed. D.E. Bloom, A. Sousa-Poza and U. Sunde, 566–577. London: Routledge.

Fishburn P.C. 1989. Retrospective on the utility theory of von Neumann and Morgenstern. *Journal of Risk and Uncertainty* **2**, 127–157.

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

Fleurbaey M. and A. Voorhoeve 2013. Decide as you would with full information! An argument against *ex ante* Pareto. In *Inequalities in Health: Concepts, Measures, and Ethics*, ed. O. Norheim, S. Hurst, N. Eyal and D. Wikler, 113–128. Oxford: Oxford University Press.

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.

Fleurbaey M. and S. Zuber 2022. A new puzzle in the social evaluation of risk. *Economics and Philosophy* **38**, 450–465.

Frankfurt H.G. 1969. Alternate possibilities and moral responsibility. *Journal of Philosophy* 23, 829–839.

Frick J. 2015. Contractualism and social risk. *Philosophy and Public Affairs* **43**, 175–223.

Fried B.H. 2003. Ex ante/ex post. *Journal of Law and Contemporary Issues* **13**, 123–160.

Gilboa I. 2009. *Theory of Decision under Uncertainty*. Cambridge: Cambridge University Press.

Greaves H. 2015. Antiprioritarianism. *Utilitas* **27**, 1–42.

Greaves H. 2017. A reconsideration of the Harsanyi–Sen–Weymark debate on utilitarianism. *Utilitas* **29**, 175–213.

Gustafsson J.E., D. Spears and S. Zuber 2023. Utilitarianism is implied by social and individual dominance. IZA Discussion Paper 16561, IZA Institute of Labor Economics. https://docs.iza.org/dp16561.pdf.

Hammond P. 1983. Ex-post optimality as a dynamically consistent objective for collective choice under uncertainty. In *Social Choice and Welfare*, ed. P.K. Pattanaik and M. Salles, 175–205. Dordrecht: Elsevier.

https://doi.org/10.1017/S0266267126100716 Published online by Cambridge University Press

Hammond P.J. 1996. Consequentialist decision theory and utilitarian ethics. In *Ethics, Rationality, and Economic Behaviour*, ed. F. Farina, F. Hahn and S. Vannucci. Oxford: Oxford University Press.
Hare C. 2011. Obligation and regret when there is no fact of the matter about what would have happened if you had not done what you did. *Noûs* **45**, 190–206.
Hare C. 2015. Statistical people and counterfactual indeterminacy. In *Identified versus Statistical Lives: An Interdisciplinary Perspective*, ed. N. Eyal, I.G. Cohen and N. Daniels, 124–136. Oxford: Oxford 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 Behaviour and Bargaining Equilibrium in Games and Social Situations*. Cambridge: Cambridge University Press.
Harsanyi J.C. 1982. Morality and the theory of rational behaviour. In *Utilitarianism and Beyond*, ed. A. Sen and B. Williams, 39–62. Cambridge: Cambridge University Press.
James A. 2012. Contractualism's (not so) slippery slope. *Legal Theory* **18**, 263–292.
Jensen K.K. 1995. Measuring the size of a benefit and its moral weight. On the significance of John Broome's interpersonal addition theorem. *Theoria* **61**, 25–60.
Kahneman D., P.P. Wakker and R. Sarin 1997. Back to Bentham? Explorations of experienced utility. *Quarterly Journal of Economics* **112**, 375–405.
Knight C. 2021. An argument for all-luck egalitarianism. *Philosophy and Public Affairs* **49**, 350–378.
Kolm S.-C. 1998. Chance and justice: social policies and the Harsanyi–Vickrey–Rawls problem. *European Economic Review* **42**, 1393–1416.
Krantz D.H., R.D. Luce, P. Suppes and A. Tversky 1971. *Foundations of Measurement, Vol. I: Additive and Polynomial Representations*. New York: Academic Press.
Lippert-Rasmussen K. 2023. Justice and bad luck. In *The Stanford Encyclopedia of Philosophy (Spring 2023 Edition)*, ed. E.N. Zalta and U. Nodelman. https://plato.stanford.edu/archives/spr2023/entries/justice–bad–luck/.
McCarthy D. 2006. Utilitarianism and prioritarianism I. *Economics and Philosophy* **22**, 335–363.
McCarthy D. 2007. Measuring life's goodness. *Philosophical Books* **48**, 303–319.
McCarthy D., K. Mikkola and J.T. Thomas 2020. Utilitarianism with and without expected utility. *Journal of Mathematical Economics* **87**, 77–113.
Mongin P. 1994. Harsanyi's aggregation theorem: multi-profile version and unsettled questions. *Social Choice and Welfare* **11**, 331–354.
Mongin P. and C. D'Aspremont 1998. Utility theory and ethics. In *Handbook of Utility Theory: Volume* **1**. *Principles*, ed. S. Barbera, P. Hammond and C. Seidl, 371–481. Dordrecht: Kluwer Academic Publishers.
Mongin P. and M. Pivato 2015. Ranking multidimensional alternatives and uncertain prospects. *Journal of Economic Theory* **157**, 146–171.
Mongin P. and M. Pivato 2016. Social evaluation under risk and uncertainty. In *The Oxford Handbook of Well-Being and Public Policy*, ed. M.D. Adler and M. Fleurbaey, 711–745. Oxford: Oxford University Press.
Muñoz D. 2025. Impartiality, anonymity, and caring who. *Free & Equal: A Journal of Ethics and Public Affairs* **1**, 472–499.
Nebel J.M. 2020. Rank-weighted utilitarianism and the veil of ignorance. *Ethics* **131**, 87–106.
Nebel J.M. 2022. Aggregation without interpersonal comparisons of well-being. *Philosophy and Phenomenological Research* **105**, 18–41.
Nebel J.M. 2024. Ethics without numbers. *Philosophy and Phenomenological Research* **108**, 289–319.
Ozbek K. 2024. Expected utility, independence, and continuity. *Theory and Decision* **97**(1), 1–22.
Parfit D. 2002. Equality or priority? In *The Ideal of Equality*, ed. M. Clayton and A. Williams, 81–125. New York: St. Martin's Press.
Pratt J.W. 1964. Risk aversion in the small and in the large. *Econometrica* **32**, 122–136.
Rakowski E. 1991. *Equal Justice*. Oxford: Oxford University Press.
Risse M. 2002. Harsanyi's 'utilitarian theorem' and utilitarianism. *Noûs* **36**, 550–577.
Roemer J.E. 1996. *Theories of Distributive Justice*. Cambridge, MA: Harvard University Press.
Scanlon T. 1998. *What We Owe to Each Other*. Cambridge, MA: Belknap Press of Harvard University Press.
Sen A. 1976. Welfare inequalities and Rawlsian axiomatics. *Theory and Decision* **7**, 243–262.

https://doi.org/10.1017/S0266267126100716 Published online by Cambridge University Press

Sen A. 1977. Non-linear social welfare functions. In *Foundational Problems in the Special Sciences. The University of Western Ontario Series in Philosophy of Science*, Vol. **10**, ed. R. Butts and J. Hintikka, 297–302. Cham: Springer.
Sen A. 2017. *Collective Choice and Social Welfare: An Expanded Edition*. Cambridge, MA: Harvard University Press.
Suppes P. 1966. Some formal models of grading principles. *Synthese* **16**, 284–306.
Thomson J.J. 1990. *The Realm of Rights*. Cambridge, MA: Harvard University Press.
VanderWeele T.J. 2017. On the promotion of human flourishing. *Proceedings of the National Academy of Sciences USA* **114**, 8148–8156.
von Neumann J. and O. Morgenstern 1944. *Theory of Games and Economic Behavior*. Princeton: Princeton University Press.
Weymark J.A. 1991. A reconsideration of the Harsanyi–Sen debate on utilitarianism. In *Interpersonal Comparisons of Well-Being*, ed. J. Elster and J.E. Roemer, 255–320. Cambridge: Cambridge University Press.

Appendix

This formal appendix lays out the model and axioms that the theorem is based upon, and then gives a proof of the theorem.

Model

- N is the finite set of n individuals.
- X is the set of final outcomes.
- $W : X \to \mathbb{R}^n$ is the well-being function.

Assumption 1. Well-being is cardinally measurable and fully comparable.
Assumption 2. The set of all well-being levels is order-isomorphic to $\mathbb{R}$.
Assumption 3. For any $(w_1, \ldots, w_n) \in \mathbb{R}^n$, there exists $x \in X$ such that $W(x) = (w_1, \ldots, w_n)$.
Notation. For all $i \in N$ and $x \in X$, $W_i(x)$ denotes i's well-being level w_i in x.

- A is the set of all probability functions on X with finite support.

Notation. For all $a \in A$, a_i is i's prospect with a.
Notation. A lottery that gives lottery a with probability α and lottery b with probability $(1 - \alpha)$ is written as $a\alpha b$.

- For all outcomes $x \in X$ and lotteries $a \in A$, if $a(x) = 1$, a and x are treated as equivalent, with corresponding abuse of notation.
- $\mathcal{R}$ is the set of all risk profiles of interest, i.e. n-tuples of rational orderings of prospects, with typical element $R = (R_1, \ldots, R_n)$.

Notation. For all $i \in N$ and $R \in \mathcal{R}$, strict preference is denoted by P_i and indifference by I_i.

- Risk-encompassing social welfare functionals are functions that return a social ordering of lotteries for each $R \in \mathcal{R}$, denoted by $\succeq_R$.

Notation. $\succ_R$ denotes strict preference, and $\sim_R$ indifference.

Axioms, Lemmas & Theorem

Rationality axioms.

Let $\succeq$ be any ordering on lotteries. Corresponding axioms are defined in the obvious way for any ordering on prospects. The following axioms are all well-known.

Continuity. For all $a, b, c \in A$, if $a \succ b \succ c$, there exist $\alpha, \beta \in (0, 1)$ such that $a\alpha c \succ b \succ a\beta c$.

For all outcomes $x \in X$ and lotteries $a \in A$, let $a[x]$ denote the probability, according to a, that an outcome at least as good as x obtains.

Stochastic Dominance. For all $a, b \in A$, if for all outcomes $x \in X$, $a[x] \geq b[x]$, then $a \succeq b$.

https://doi.org/10.1017/S0266267126100716 Published online by Cambridge University Press

Independence. For all $a, b, c \in A$ and every $\alpha \in (0,1)$, if $a \succeq b$ then $a\alpha c \succeq b\alpha c$.

The Allais preferences (Allais 1953) are regimented as violations of T-Independence (Ozbek 2024: 5), which is entailed by Independence.

T-Independence. For all $a, b, c, d \in A$ and every $\alpha \in (0,1)$ if $a\alpha c \succeq b\alpha c$ then $a\alpha d \succeq b\alpha d$.

We next note two more entailments of Independence (adapting Ozbek 2024: 5–6).

S-Independence. For all $a, b \in A$ and $\alpha \in (0,1)$, if $x_* \in X$ is such that for all outcomes y in the support of either a or b, $y \succeq x_*$, then $a \succeq b$ entails $a\alpha x_* \succeq b\alpha x_*$.

Betweenness. For all $a, b \in A$ and $\alpha \in (0,1)$, if $a \succeq b$, then $a \succeq a\alpha b \succeq b$.

Lemma 1 (adapting Ozbek 2024: 6, Proposition 1). *T-Independence and S-Independence, and T-Independence and Betweenness, each entail Independence.*

Proof. The proof follows Ozbek (2024: 16–17) with minor adaptations, and is omitted.

We say that the **Allais-Compatible Package** is satisfied if and only if: for all $R \in \mathcal{R}$, each individual entry of R satisfies Continuity and either S-Independence or Betweenness, and $\succeq_R$ satisfies Continuity and Stochastic Dominance.

Ethical axioms.

Strong Pareto. For all profiles $R \in \mathcal{R}$ and lotteries $a, b \in A$, if $a_i R_i b_i$ for all i, then $a \succeq_R b$, and if, moreover, $a_i P_i b_i$ for some i, then $a \succ_R b$.

Interprofile Anonymity. For all $a, b, c, d \in A$, and $R, R' \in \mathcal{R}$, if there exists a permutation σ of N such that for all $i \in N$, $a_i = b_{\sigma(i)}$ and $c_i = d_{\sigma(i)}$, and for all concerned individuals i, $R_i = R'_{\sigma(i)}$, then $a \succeq_R c \iff b \succeq_{R'} d$.

Lemma 2. *Strong Pareto, Interprofile Anonymity and the Allais-Compatible Package entail expected utility theory for individual and social preference.*

Proof. We derive Independence from Strong Pareto, Interprofile Anonymity, and the Allais-Compatible Package (adapting Nebel 2020 and Bradley 2022 to the framework of lotteries; note that we don't need Outcome Anonymity, which these authors appeal to).

Suppose that individual preferences need not satisfy T-Independence. Then for some $R \in \mathcal{R}$, $R_i = R_j$ for all $i, j \in N$, and there exist prospects p, q, r, s and $\alpha \in (0,1)$ such that $p\alpha s R_i q\alpha s$ but $q\alpha r P_i p\alpha r$ for all $i \in N$. Fix $a, b \in A$ as follows: $a = \underline{a}\alpha c$ and $b = \underline{b}\alpha c$, with everyone but two individuals, j and k, receiving some arbitrary well-being level w for sure, and $\underline{a}_j = p$, $\underline{a}_k = q$, $\underline{b}_j = q$ and $\underline{b}_k = p$, $c_j = s$, $c_k = r$. In particular, for all well-being levels w_p, w_q in the support of p and q respectively, choose $x \in X$ such that $W_j(x) = w_p$, $W_k(x) = w_q$, and let $\underline{a}(x) = p(w_p)q(w_q)$, and choose $y \in X$ such that $W_j(y) = w_q$, $W_k(y) = w_p$, and let $\underline{b}(y) = p(w_p)q(w_q)$. Note that there exists a bijection f from outcomes in the support of $\underline{a}$ to outcomes in the support of $\underline{b}$ such that $\underline{a}(x) = \underline{b}(f(x))$ and $W_j(x) = W_k(f(x))$, $W_k(x) = W_j(f(x))$; by Interprofile Anonymity, $x \sim_R f(x)$. Given this construction of a and b, Stochastic Dominance for social preference entails $a \sim_R b$. By Strong Pareto, however, $a \succ_R b$. Contradiction. Therefore, individual preferences satisfy T-Independence for all $R \in \mathcal{R}$, and hence Independence by Lemma 1. Since they also satisfy Continuity and are orderings, the vNM theorem applies and there exist expected utility representations for each entry R_i of any $R \in \mathcal{R}$ (see e.g. Gilboa 2009: 79–86). Moreover, since individual preferences match well-being levels, each individual's vNM utility function is increasing in well-being.

Suppose next that social preference need not satisfy Independence. Then for some $R \in \mathcal{R}$ there exist $a, b, c \in A$ and $\alpha \in (0,1)$ such that $a \succeq_R b$ but $f = b\alpha c \succ_R e = a\alpha c$. Define a^*, b^*, c^* as follows: for all $i \in N$ a_i^* (respectively b_i^*, c_i^*) is the certainty equivalent of a_i (respectively b_i, c_i). By Strong Pareto $a^* \sim_R a$, $b^* \sim_R b$; hence $a^* \succeq_R b^*$. Let $e^* = a^*\alpha c^*$ and $f^* = b^*\alpha c^*$. By Stochastic Dominance for social preference, $e^* \succeq_R f^*$. By Strong Pareto, $e^* \sim_R e$, $f^* \sim_R f$; hence $e \succeq_R f$. Contradiction. Therefore, social preference satisfies Independence for all $R \in \mathcal{R}$. Since each output social preference is an ordering that also satisfies Continuity, the vNM theorem applies and for all $R \in \mathcal{R}$, there exist an expected utility representation for $\succeq_R$.

Theorem. *Assume the Allais-Compatible Package. Then a risk-encompassing social welfare functional satisfies Strong Pareto and Interprofile Anonymity if and only if it is Impartial Risk-Prioritarian.*

Proof. ($\Rightarrow$) Given Lemma 2, Harsanyi's (1955) social aggregation theorem then allows us to infer that for all $R \in \mathcal{R}$, we can choose expected utility representations $U_R, U_{R_1}, \ldots, U_{R_n}$ for social and individual

https://doi.org/10.1017/S0266267126100716 Published online by Cambridge University Press

	1	2	others
e	w^+	w	w
f	w	w^+	w

Figure 3. Lotteries e and f defined.

	i	j	k	l	others
R	$\star$	R_j	$\circ$	R_l	R_{others}
R'	R'_i	$\star$	R'_k	$\circ$	R'_{others}
R''	$\star$	R'_i	$\circ$	R'_k	R'_{others}

Figure 4. Depiction of profiles appealed to in the proof.

preferences respectively, such that for all $a \in A$, $U_R(a) = U_{R_1}(a_1) + \ldots + U_{R_n}(a_n)$ (see Broome 1991 for this way of stating Harsanyi's theorem). We want to show the following:

(i) for all $R \in \mathcal{R}$, and $i, j \in N$, if $R_i = R_j$, then $U_{R_i} = U_{R_j} + \beta$ (this means: within a profile, the well-being of individuals with the same risk attitude is weighted in the same way);
(ii) for all $R, R' \in \mathcal{R}$, and $i, j, k, l \in N$, if $R_i = R'_j$ and $R_k = R'_l$ then $U_{R_i} = \alpha U_{R'_j} + \beta_1$ and $U_{R_k} = \alpha U_{R'_l} + \beta_2$, with constant $\alpha > 0$ and $\beta_1, \beta_2 \in \mathbb{R}$ (this means: across profiles, the well-being of individuals with the same risk attitudes is weighted in the same way).

We first prove that if (i) fails, then Interprofile Anonymity is violated. By contraposition, we then get that if Interprofile Anonymity holds, (i) also does. So, we fix $R \in \mathcal{R}$, and let $R_1 = R_2$ but $U_{R_1} = \alpha U_{R_2} + \beta$ for $1, 2 \in N$, with $\alpha \neq 1$. Without loss of generality, assume $\alpha > 1$. Then consider the two lotteries $e, f \in A$ in Figure 3, where $w^+ > w \in \mathbb{R}$.

Lottery $e \in A$ results for sure in everyone but 1 at well-being w, and 1 at $w^+ > w$. Lottery $f \in A$ results for sure in everyone but 2 at well-being w, and 2 at w^+. Then $e \succ_R f$, which violates Interprofile Anonymity. To see this, let $R = R'$ and $a = d, b = c$ in the statement of the principle. Then take the permutation $\sigma(1) = 2, \sigma(2) = 1$, and $\sigma(k) = k$ for all $k \neq 1, 2 \in N$. For all individuals $i \in N$, we have $e_i = f_{\sigma(i)}$, and $f_i = e_{\sigma(i)}$, and for all concerned individuals, i.e. 1 and 2, we have $R_1 = R_{\sigma(1)=2}$ and $R_2 = R_{\sigma(2)=1}$. But it's not the case that $e \succeq_R f$ iff $f \succeq_R e$ – since $e \succ_R f$. By contraposition, if Interprofile Anonymity holds, then (i) follows.

We show (ii) by contradiction. Fix $R, R' \in \mathcal{R}$, and $i, j, k, l \in N$ with $R_i = R'_j = *$ and $R_k = R'_l = \circ$, but suppose for contradiction that $U_{R_i} = \alpha_1 U_{R'_j} + \beta_1$ and $U_{R_k} = \alpha_2 U_{R'_l} + \beta_2$, with $\alpha_1 \neq \alpha_2$. Without loss of generality, we assume that $\alpha_1 > \alpha_2$. Figure 4 depicts R and R', and another profile R'' that we will need to consider.

Arbitrarily select $w, w^+ \in \mathbb{R}$, where $w^+ > w$. Then, for $U_{R'_j}, U_{R'_l}, U_{R_i}, U_{R_k}$ the difference in utility between (degenerate prospects yielding for sure) w^+ and w is positive, since all individual utility functions are strictly increasing in well-being. Let ϵ be some positive real number strictly smaller than any of these four differences in utility, and M some other positive real number strictly greater than any of these four differences in utility. Next, consider prospect v, which gives to the individual who receives it a p probability of w^+, and a $(1-p)$ probability of w, with $p = \frac{\varepsilon}{2M}$. Notice that for all of $U_{R'_j}, U_{R'_l}, U_{R_i}, U_{R_k}$, the difference in utility between v and w is positive but strictly smaller than ϵ. This can be shown as follows, with U standing in for any one of the four utility functions under consideration. Since U is expectational,

$$U(v) = pU(w^+) + (1-p)U(w)$$

$$U(v) = pU(w^+) + U(w) - pU(w)$$

$$U(v) - U(w) = p[U(w^+) - U(w)]$$

$$U(v) - U(w) = \frac{\varepsilon}{2M}[U(w^+) - U(w)] < \frac{\varepsilon}{2M}M = \frac{\varepsilon}{2} < \varepsilon$$

Look next at Figure 5, illustrating four lotteries $e, f, g, h \in A$.

https://doi.org/10.1017/S0266267126100716 Published online by Cambridge University Press

	i	j	k	l	others
e	v	w	w	w	w
f	w	v	w	w	w
g	w	w	v	w	w
h	w	w	w	v	w

Figure 5. Lotteries e, f, g, h defined.

	i	j	k	l	others
e	v	w	w	w	w
f	w	v	w	w	w
$g*$	w	w	v^+	w	w
$h*$	w	w	w	v^+	w

Figure 6. Lotteries $g*, h*$ defined; e, f also depicted for comparison.

Consider the permutation σ of N such that $\sigma(i) = j, \sigma(j) = i, \sigma(k) = l, \sigma(l) = k$, and $\sigma(m) = m$ for $m \neq i, j, k, l \in N$. Note that $g_i = h_{\sigma(i)}$ and $e_i = f_{\sigma(i)}$, and similarly for all other individuals. Also, $R_i^{''} = R'_{\sigma(i)}$, and similarly for the remaining concerned individuals j, k, l. By Interprofile Anonymity, $g \succeq_{R''} e$ iff $h \succeq_{R'} f$. Now look at R and $R^{''}$, and consider the identity permutation. By Interprofile Anonymity, $g \succeq_R e$ iff $g \succeq_{R''} e$. This can be seen by letting $a = b$ and $c = d$ in the statement of the principle. It then follows that $g \succeq_R e$ iff $h \succeq_{R'} f$. There are three cases. If $h \sim_{R'} f$, since $\alpha_1 > \alpha_2$, it follows tht $e \succ_R g$, which contradicts the verdict just reached using Interprofile Anonymity twice. If $f \succ_{R'} h$, we need an additional step. First notice that $f \succ_{R'} h$ entails that $U_{R'_j}(v) - U_{R'_j}(w) > U_{R'_l}(v) - U_{R'_l}(w)$. Define next v^+ as a prospect that gives to the individual who receives it a q probability of w^+ and a $(1 - q)$ probability of w, with

$$q = \frac{U_{R'_j}(v) - U_{R'_j}(w)}{U_{R'_l}(w^+) - U_{R'_l}(w)}$$

Notice that $0 < q < 1$, since $0 < U_{R'_j}(v) - U_{R'_j}(w) < \epsilon < U_{R'_l}(w^+) - U_{R'_l}(w)$. We now show that $U_{R'_l}(v^+) - U_{R'_l}(w) = U_{R'_j}(v) - U_{R'_j}(w)$. Since l is an expected utility maximizer,

$$U_{R'_l}(v^+) = qU_{R'_l}(w^+) + (1 - q)U_{R'_l}(w)$$

Rearranging,

$$U_{R'_l}(v^+) - U_{R'_l}(w) = q\left[U_{R'_l}(w^+) - U_{R'_l}(w)\right]$$

So,

$$U_{R'_l}(v^+) - U_{R'_l}(w) = \frac{U_{R'_j}(v) - U_{R'_j}(w)}{U_{R'_l}(w^+) - U_{R'_l}(w)}\left[U_{R'_l}(w^+) - U_{R'_l}(w)\right] = U_{R'_j}(v) - U_{R'_j}(w)$$

With $g*, h* \in A$ as defined in Figure 6, we have $h* \sim_{R'} f$ but $e \succ_R g*$ (since $\alpha_1 > \alpha_2$), which can be shown to contradict Interprofile Anonymity.

To see the contradiction, consider again first profile R'' and R' as above, and note that Interprofile Anonymity entails $g* \succeq_{R''} e$ iff $h* \succeq_{R'} f$. Next, consider R'' and R, and note that Interprofile Anonymity entails $g* \succeq_R e$ iff $g* \succeq_{R''} e$. Therefore, $g* \succeq_R e$ iff $h* \succeq_{R'} f$. Finally, if $h \succ_{R'} f$, we need an analogous additional step. Define $v^\dagger$ as a prospect that gives to the individual who receives it an r probability of w^+ and a $(1 - r)$ probability of w, with

$$r = \frac{U_{R'_l}(v) - U_{R'_l}(w)}{U_{R'_j}(w^+) - U_{R'_j}(w)}$$

Notice that $0 < r < 1$. We show that $U_{R'_j}(v^\dagger) - U_{R'_j}(w) = U_{R'_l}(v) - U_{R'_l}(w)$. Since j is an expected utility maximizer,

https://doi.org/10.1017/S0266267126100716 Published online by Cambridge University Press

	i	j	k	l	others
$e*$	$v^{\dagger}$	w	w	w	w
$f*$	w	$v^{\dagger}$	w	w	w
g	w	w	v	w	w
h	w	w	w	v	w

Figure 7. Lotteries $e*, f*$ defined; g, h also depicted for comparison.

$$U_{R'_j}\left(v^{\dagger}\right) = rU_{R'_j}(w^{+}) + (1-r)U_{R'_j}(w)$$

Rearranging,

$$U_{R'_j}\left(v^{\dagger}\right) - U_{R'_j}(w) = r[U_{R'_j}(w^{+}) - U_{R'_j}(w)]$$

So,

$$U_{R'_j}\left(v^{\dagger}\right) - U_{R'_j}(w) = \frac{U_{R'_l}(v) - U_{R'_l}(w)}{U_{R'_j}(w^{+}) - U_{R'_j}(w)}\left[U_{R'_j}(w^{+}) - U_{R'_j}(w)\right] = U_{R'_l}(v) - U_{R'_l}(w)$$

With $e*, f* \in A$ as given in Figure 7, we have $h \sim_{R'} f*$ but $e* \succ_R g$, which again can be shown to contradict Interprofile Anonymity by similar reasoning as above.

($\Leftarrow$) This direction is immediate, and the proof is omitted.

Pietro Cibinel is an Assistant Professor in the Department of Philosophy at the University of North Carolina, Chapel Hill. He works at the intersection of ethics and decision theory, with his research investigating key connections between rational choice, social choice and welfare measurement.

Cite this article: Cibinel P. The Risk-Priority View. *Economics and Philosophy*. https://doi.org/10.1017/S0266267126100716

https://doi.org/10.1017/S0266267126100716 Published online by Cambridge University Press