Welfare vs. Utility

Draft[1]
November 2022 / June 2024
Franz Dietrich

Abstract

Ever since the Harsanyi-Sen debate, it is controversial whether someone's welfare should be measured by her von-Neumann-Morgenstern (VNM) utility, for instance when analysing welfare intensity, social welfare, interpersonal welfare comparisons, or welfare inequality. As we show, natural working assumptions lead to a different welfare measure, which addresses familiar concerns about VNM utility while also requiring only ordinal evidence, such as observed choices or self-reported comparisons. Using this measure instead of VNM utility has different implications, for instance for social welfare and policy recommendations, in riskless or risky contexts. VNM utility is shown to be have two determinants, namely welfare and attitude to risk in welfare.

1 Introduction

How should someone's welfare be measured? This is an important methodological question. In practice, economists often use VNM utility to measure welfare. The reason is that a VNM utility function (if existent) rests solely on ordinal evidence, such as revealed preferences or self-reported comparisons between lotteries. The key limitation of such a welfare measure is that it arguably fails to adequately capture non-ordinal information like information about welfare intensity or absolute welfare levels. Non-ordinal welfare information is however essential for many applications, such as: aggregating individual into social welfare, comparing welfare levels (or differences) across people, measuring inequality in welfare, and making welfare-based policy recommendations. The question of whether VNM utility captures non-ordinal welfare information (and is thus useful in such applications) is however controversial. It has culminated in the Harsanyi-Sen debate in the 1970s, and counts today among the most notorious open foundational problems in welfare economics and formal ethics.

Critics of VNM utility as a welfare measure have so far not come up with an alternative measure that is also based on ordinal evidence. The lack of ordinal foundations exposes these critics to the 'non-observable' objection, if one follows the ordinalist tradition according to which all evidence about welfare is ordinal. The ordinalist notion of evidence is itself controversial, but we will make this classic economic assumption here.

Rather than settling the Harsany-Sen debate substantively, this paper provides a proof of concept for the VNM-sceptic position, by showing that purely ordinal

[1] Acknowledgements TBA. This work has been presented at SEAT (Paris, July 2023), Centre d'Economie de la Sorbonne (Paris 1, November 2023), and the Workshop on the Foundations of Decision Theory (LMU Munich, June 2024).

evidence leads to a *different* welfare measure if one accepts some plausible working assumptions. This measure will respond to classic objections against VNM utility as a welfare measure.

We start from the familiar idea, analysed notably in Bell and Raiffa (1988), that someone's VNM utility function is affected by two different things: the welfare or 'intrinsic utility' derived from outcomes, and the attitude to risk in welfare or 'intrinsic risk'. In the example in Figure 1, an individual has a concave VNM utility

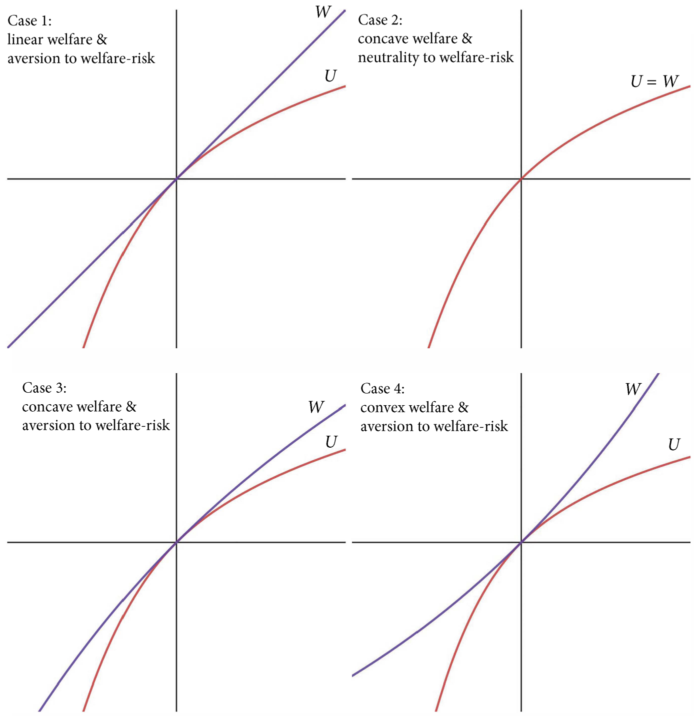

Figure 1: A VNM utility function and four possible explanations in terms of welfare and intrinsic risk attitude

function over possible wealth levels (plotted in red), and we consider four alternative explanations, which differ in the agent's welfare function (plotted in blue) and her intrinsic risk attitude:

- *Case 1: constant marginal welfare & intrinsic risk aversion.* The welfare from an extra 1$ is the same regardless of initial wealth, and facing a lottery of wealth levels is worse than having the expected welfare for sure (we say 'expected welfare' rather than 'expected wealth' to refer to the attitude to risk in welfare rather than in wealth). Welfare is linear, and utility is concave in welfare, thus concave in wealth.

- *Case 2: diminishing marginal welfare & intrinsic risk neutrality.* An extra $1 gives less extra welfare if the person is wealthier, and facing a lottery is indifferent to having the expected welfare for sure. Welfare is concave, and utility coincides with welfare.

- *Case 3: diminishing marginal welfare & intrinsic risk aversion.* Welfare is concave, and utility is concave in welfare and thus (more) concave in wealth.
- *Case 4: increasing marginal welfare & strong intrinsic risk aversion.* An extra $1 gives *more* extra welfare if the person is wealthier (implausibly), and a lottery is much worse than having the expected welfare for sure. Welfare is convex, and utility is so strongly concave in welfare that it is concave in wealth.

In all four plots in Figure 1, the utility and welfare functions are normalised so that they take value 0 and derivative 1 at a fixed wealth level.

The problem is that it is impossible on the standard approach to know the agent's welfare function based on the utility function: none of the four cases can be ruled out empirically. In result, one cannot assess social welfare, measure welfare inequality, or make welfare-driven policy choices. This paper will propose a way to overcome this dilemma. It will make welfare indirectly observable.

Our premise that VNM utility has two determinants – welfare (intrinsic utility) and attitude to welfare-risk (intrinsic risk attitude) – will resonate with most rational choice theorists, who indeed routinely invoke both determinants. Still, this view can be challenged in different ways. Let us now sketch some prominent understandings of 'utility' and 'welfare', only some of which are compatible with our picture.

Sen (1977) and Weymark (1991) forcefully argue for distinguishing between someone's welfare with her VNM utility. Bell and Raiffa (1988), Nissan-Rosen (2015), Dietrich and Jabarian (2022) and many others endorse the distinction. Yet for instance Harsanyi, Broome (1991), McCarthy, Greaves (2017), and McCarthy et al. (2020) question or even reject the distinction. Fleurbaey and Mongin (2016) take a nuanced view.

For John Harsanyi, welfare just *is* VNM utility. Still his notion of welfare is not simply ordinal: for him, welfare is not just a numerical representation of ordinal comparisons, but (when suitably normalised) contains information about welfare intensity and interpersonal comparisons. On his view, *ordinal* evidence (i.e., preferences) generate a *non-ordinal* welfare measure. Unlike us, he abandons the distinction between welfare-related and risk-attitudinal aspects: for him, risk aversion comes precisely from diminishing marginal welfare. To us, this reduction conflates two entirely distinct phenomena. It is perfectly possible to dislike unpredictability without having diminishing marginal welfare, or vice versa.

First-generation economists used to perfectly distinguish between someone's welfare and her taste for risk. They used the term 'utility' in a sense that is free of risk preferences and corresponds to what we call 'welfare' or 'intrinsic utility' rather than 'VNM utility'. Accordingly, their 'law of diminishing marginal utility' does not reflect any risk aversion but diminishing benefits from consumption. Yet they did not tell us how to measure welfare from ordinal data, a problem tackled here.

Arrow and Pratt, the fathers of the modern theory of risk aversion (Pratt 1964, Arrow 1965), regard risk preferences as the sole origin of the VNM utility function – at least following the dominant perception of their theory and the label 'theory of *risk* aversion'. Their theory could be reinterpreted more broadly, as a theory about the combination of (intrinsic) risk attitude and marginal welfare. On this reinterpretation, VNM utility is a hybrid quantity combining welfare and (intrinsic) risk attitudes, in line with our approach. Yet, although Arrow-Pratt's theory (so re-interpreted) is compatible with an independent notion of welfare, it does not tell us how to derive it.

Peter Wakker (2010) takes an interestingly different view. He defends a clean separation between risk attitude and welfare, but takes only welfare to affect VNM utility. The risk attitude instead affects the weighing of subjective probabilities within a choice model that follows prospect theory and rank-dependence rather than ordinary expected-utility theory – an approach similar to Buchak's (2013). On this view, VNM utility has a single determinant – marginal welfare – yet not because the risk attitude is merely a by-product of marginal welfare (as for Harsanyi) but because the risk attitude enters at a different level (the probability weighting). Given that VNM utility is then a purely welfare-theoretic construct free of the risk attitude, Arrow-Pratt's theory 'of risk aversion' would be reinterpreted as a theory 'of diminishing marginal welfare'. And our question of how to measure welfare would have an easy answer: use VNM utility.

Finally, welfare measures have been studied extensively using *measurement theory*, under names such as 'measuring strength-of-preference' or 'measuring preference-intensity'. See in particular Krantz et al. (1971), Shapley (1975), Basu (1982), and for particularly general results Wakker (1988, 1989), Köbberling (2006) and Pivato (2013). This literature pursues an interestingly different agenda, as the evidential basis is not an order over alternatives, but an order R over alternative *pairs*, called a 'difference order', where $(x, y)R(x', y')$ means 'a change from x to y is at least as good as a change from x' to y''. A welfare measure is then a representation of this difference order.[2] Like a VNM utility function, such a welfare measure is typically unique up to increasing affine transformation. But, unlike a VNM function (standardly interpreted), it captures welfare *intensity*, not just welfare comparisons, and it is derived from non-choice data, since the difference order R is not revealed by choice. By contrast, we aim to measure welfare based on purely ordinal observations, where 'ordinal' for us refers to a binary order over options, not a difference order.

2 A partially unique welfare measure

We fix a set X of *situations*, in which the welfare of a given individual is to be measured. Although we could work without any assumptions on X (as shown in the appendix), the main text takes situations to be real numbers or more generally vectors of real numbers. Typical real-valued situations are wealth levels, health levels, or consumption index levels. Typical vector-valued situations are consumption bundles, vectors of functionings (Sen 1985), or wealth-health-education triples. Technically, the main text lets X be a non-empty open connected subset of $\mathbb{R}^k$ for some $k \geq 1$, e.g., $\mathbb{R}^k$, $(0, \infty)^k$ or $(0, 1)^k$. Readers can focus on the base-line case that $k = 1$, in which X is a non-empty open interval, e.g., $\mathbb{R}$ or $(0, \infty)$ or $(0, 1)$.

A *welfare (or intrinsic utility) measure* is a function $W : X \to \mathbb{R}$, where $W(x)$ represents the person's welfare at x, or, under alternative interpretations that we set aside, the intensity of preference for x or the intrinsic value attached to x. Welfare is not directly observable. Instead we observe ordinal comparisons, more precisely

[2] A function W 'represents' a difference order R over a set of alternatives if $W(x) - W(y) \geq W(x') - W(y')$ for all alternatives.

comparisons of risky prospects, representing actions or policies with unknown outcome. Technically, a *prospect* is a lottery over X with finite support. Let $\mathcal{P}$ be the set of these prospects. We have $X \subseteq \mathcal{P}$, by identifying any situation x in X with the riskless prospect where x occurs for sure. Our observable primitive is a binary relation $\succeq$ on $\mathcal{P}$, called an *observed order*, where '$x \succeq y$' means that x is observably at least as good as y for the person. Let $\succ$ and $\sim$ denote the corresponding strict and asymmetric relations, where '$x \succ y$' and '$x \sim y$' mean that x is observably better than resp. as good as y for the person.[3] The source of observation might consist in choice behaviour, reported self-assessments, third-party assessments, or perhaps neurophysiological data.

The classic economic move would be to identify welfare with VNM utility. A *VNM utility representation* of $\succeq$ is a function $U : X \to \mathbb{R}$ such that the prospects are ranked by expected utility, i.e., for all prospects $p, q \in \mathcal{P}$, $p \succeq q$ if and only if $\mathbb{E}_p(U) \geq \mathbb{E}_q(U)$. If existent, such a representation is unique up to increasing affine transformation. We will not use a VNM representation to measure welfare, for reasons discussed above.

We now introduce a first hypothesis about the welfare measure W in relation to the observable $\succeq$. To motivate it, recall the following classic condition, where a *certainty equivalent* of a prospect $p \in \mathcal{P}$ is a situation $x_p \in X$ such that $p \sim x_p$:

Constant Absolute Risk Aversion (CARA), defined for $X \subseteq \mathbb{R}$: If all outcomes of a prospect increase by a fixed amount, then the certainty equivalent increases by this amount. Formally, for all $\Delta > 0$ and all prospects $p, q \in \mathcal{P}$ with a certainty equivalent x_p resp. x_q, if $p(x) = q(x + \Delta)$ for all $x \in \mathbb{R}$, then $x_q = x_p + \Delta$.[4]

CARA is widely empirically violated, in favour of decreasing rather than constant absolute risk aversion (Chiappori and Paiella 2011). The fundamental problem is arguably that, as soon as marginal *welfare* is diminishing, risk aversion relative to *outcomes* tends to be decreasing, not constant. Indeed, if a risky wealth prospect is translated upwards, then it moves into a region of higher wealth and thus lower marginal welfare, so that the new prospect contains less risk *in welfare*, i.e., less risk in a subjectively relevant sense. A 50-50 lottery between wealth \$0 and wealth \$1,000,000 contains huge risk in welfare: $W(0) \ll W(1,000,000)$. But the translated 50-50 lottery between wealth \$10.000.000 and \$11.000.000 contains almost no risk in welfare: $W(10,000,000) \approx W(11,000,000)$. The first lottery should thus be equivalent to a wealth level close to the worse outcome of \$0, but the second lottery to a wealth level close to the average outcome of $\$10,500,000$, violating CARA.

Economists usually react to the empirical violations of CARA by replacing CARA with some condition allowing for decreasing absolute risk aversion, for instance the condition of hyperbolic absolute risk aversion (HARA). Our response is different. The fundamental problem with CARA and its classic alternatives lies in the choice of outcomes rather than welfare as the 'currency' or 'level of description' of risk

[3]For all $p, q \in \mathcal{P}$, $p \succ q$ if and only if $p \succeq q$ and not $q \succeq p$, and $p \sim q$ if and only if $p \succeq q$ and $q \succeq p$.

[4]To make '$p(x)$' and '$q(x + \Delta)$' well-defined even if x resp. $x + \Delta$ fall outside X, we identify any lottery over X ($\subseteq \mathbb{R}$) with its extension to $\mathbb{R}$, which is zero within $\mathbb{R} \backslash X$. Equivalently to CARA, risk premia are invariant to translating prospects by a fixed amount (assuming certainty equivalents are unique). Here, the *risk premium* of a prospect $p \in \mathcal{P}$ with a (unique) certainty equivalent is the gap $\overline{p} - x_p$ between p's expectation $\overline{p} = \sum_{x \in X} p(x)$ and certainty equivalent x_p.

aversion. More naturally, risk aversion would be described as aversion to risk in welfare, since the 'true' risk comes from the possibility of *differently good* outcomes, not just *different* outcomes. We thus replace CARA with the following condition, where an *equivalent welfare* of a prospect $p \in \mathcal{P}$ is a welfare level $w_p = W(x_p)$ that is achieved in a situation $x_p \in X$ such that $x_p \sim p$:

Constant Intrinsic Risk Aversion (CIRA): If all *welfare* outcomes of a prospect increase by a fixed amount, then the equivalent *welfare* increases by this amount. Formally, for all $\Delta > 0$ and all prospects $p, q \in \mathcal{P}$ with an equivalent welfare w_p resp. w_q, if $p(W = w) = q(W = w + \Delta)$ for all $w \in \mathbb{R}$, then $w_q = w_p + \Delta$.[5]

CIRA requires a coherent or stable attitude to intrinsic risk, i.e., risk in resulting welfare. For instance, if at a low welfare of 1 the agent likes a 50:50 gamble of gaining 2 units or losing 1 unit *of welfare*, then she still likes this gamble when starting at a welfare of 50 or 100. CIRA provides an attractive explanation for the empirical finding of decreasing aversion to risk *in outcomes*, assuming diminishing marginal welfare.

Why do we replace CARA with CIRA rather than with some classic assumption of decreasing absolute risk aversion? For one, CIRA is no longer a condition on preferences alone, but one on the relation between preferences $\succeq$ and welfare W. This is precisely how it should be given out aim to make W observable. Classic authors instead aim at representing and predicting choices, and for that reason avoid conditions with extra parameters such as W. A second reason not to use a standard alternative to CARA is that any given alternative (such as HARA) is only plausible for a highly special and systematic shape of the welfare function W. The condition will fail if W behaves unsystematically, for instance is strongly convex in some ares and less convex or linear in other parts. But there are no general grounds to exclude unsystematic behaviour of welfare. There is nothing irrational or surprising about an unsystematic W, since W is not subject to rationality but partly to physiology. The plausibility of CIRA does not hinge on any particular shape of the welfare function, which could by arbitrarily 'crazy'.

Still, CIRA is debatable, as discussed in Section 5, where we present a generalised condition and theorem.

Before introducing further hypotheses, let us pause and see what CIRA implies on its own. We shall restrict attention to well-behaved welfare measures and observed orders $\succeq$. We call a function $W : X \to \mathbb{R}$ *well-behaved* (relative to $\succeq$) if it is

- *compatible with riskless comparisons*: for any situations $x, y \in X$, $W(x) \geq W(y) \Leftrightarrow x \succeq y$. That is, W is ordinally equivalent to the restriction of $\succeq$ to the set X of riskless prospects.
- *regular*: W is smooth with nowhere zero derivative W'.[6]

[5] Equivalently, *intrinsic* risk premia are invariant to translating prospects by a fixed amount (assuming a prospect's equivalent welfare is unique). Here, the *intrinsic risk premium* of a prospect $p \in \mathcal{P}$ with a (unique) equivalent welfare w_p is the gap $\mathbb{E}_p(W) - w_p$ between p's expected and equivalent welfare.

[6] 'Smooth' means that W is differentiable arbitrarily many often. In the muli-dimensional case $X \subseteq \mathbb{R}^k$ with $k \geq 2$, W' is of course a vector $(\frac{d}{dx_1}W, ..., \frac{d}{dx_k}W)$. It is 'nowhere zero' if it at each $x \in X$ it is not the zero vector, i.e., at least one partial derivative is non-zero.

For instance, if X is a set of wealth levels $X \subseteq \mathbb{R}$ and if $x > y \Rightarrow x \succ y$ ('more wealth is better'), then well-behaved welfare measures are smooth functions $W : X \to \mathbb{R}$ with $W' > 0$.

The observed order $\succeq$ is *well-behaved* if it has a VNM representation U that is regular (and thus automatically well-behaved[7]).

Proposition 1 *Given a well-behaved observed order $\succeq$, a well-behaved welfare measure $W : X \to \mathbb{R}$ satisfies CIRA if and only if*

$$W = \log(\rho U + 1)/\rho$$

for some VNM representation U of $\succeq$ and some $\rho \in \mathbb{R}$ (called the 'intrinsic risk proneness') such that $\rho U + 1 > 0$.

If $\rho = 0$, then '$\log(\rho U + 1)/\rho$' stands for U $(= \lim_{\rho\to 0} \log(\rho U + 1)/\rho)$. The value of ρ is not free: it must ensure that $\rho U + 1 > 0$. For instance, $\rho = 0$ if $\sup U = \infty$ and $\inf U = -\infty$. Interpretively, ρ measures the attitude to *intrinsic* risk, that is, risk in resulting welfare:

- if $\rho > 0$, the agent is strictly intrinsic risk prone, as VNM utility is convex in welfare,
- if $\rho < 0$, the agent is intrinsic risk averse, as VNM utility is concave in welfare,
- if $\rho = 0$, the agent is intrinsic risk neutral, as VNM utility is linear in welfare.

In case of a weak inequality $\rho \geq 0$ resp. $\rho \leq 0$ the agent is *weakly* intrinsic risk prone resp. averse. Although the exact value of ρ is usually underdetermined, the sign of ρ – and hence the qualitative intrinsic risk attitude – is often determined. Indeed, because ρ must satisfy $\rho U + 1 > 0$, the agent is necessarily

- intrinsic risk neutral if $\sup U = \infty$ and $\inf U = -\infty$, as then $\rho = 0$,
- weakly intrinsic risk prone if $\sup U = \infty$ and $\inf U \neq -\infty$, as then $\rho \geq 0$,
- weakly intrinsic risk averse if $\sup U \neq \infty$ and $\inf U = -\infty$, as then $\rho \leq 0$.

The next section will formally justify our claims about risk attitude.

3 Explaining standard utility and risk attitude by intrinsic utility and intrinsic risk attitude

This section explores the structure of classic VNM utility and Arrow-Pratt risk attitude, by decomposing both quantities into their two influences, welfare and intrinsic risk attitude. These decompositions show that classic utility and classical risk attitude are hybrid constructs that reflect an interplay of welfare and intrinsic risk attitude. At this stage the decomposition will still be empirically underdetermined,

[7] Any VNM representation is compatible with riskless comparisons.

because welfare is not yet uniquely identified. Unique identification will be achieved later.

The classic economic theory of risk attitude after Arrow and Pratt measures the risk attitude as follows, by assuming the one-dimensional case $X \subseteq \mathbb{R}$:

Definition 1 *The* classical (or Arrow-Pratt) risk proneness *of a well-behaved observed order $\succeq$, for $X \subseteq \mathbb{R}$, is the (well-defined[8]) function $\rho_{AP} = \frac{U''}{U'}$ where U is any VNM representation of $\succeq$. If ρ_{AP} is constant, it is identified with its single value.*

This measure is sometimes re-defined relative to a different quantity than the outcome in X. We here re-define it relative to an unobservable quantity, namely welfare, in order to eliminate any distortion of the risk attitude by welfare- rather than risk-related factors.

Definition 2 *The* intrinsic risk proneness *of a well-behaved observed order $\succeq$ w.r.t. a well-behaved welfare measure W is the (well-defined[9]) function $\rho_W = \frac{d^2U/dW^2}{dU/dW}$ where U is any VNM representation of $\succeq$. If ρ_W is constant, it is identified with its single value.*

We can now formally confirm our earlier interpretation of the parameter 'ρ' of our welfare measure:

Remark 1 *In Proposition 1, the intrinsic risk proneness ρ_W is constant and equals the parameter ρ in the welfare measure $W = \log(\rho U + 1)/\rho$.*

VNM utility can be decomposed into its two determinants, welfare (intrinsic utility) and intrinsic risk proneness, by simply solving the equation '$W = \log(\rho U + 1)/\rho$' for U and replacing ρ with ρ_W:

Corollary 1 *A well-behaved observed order $\succeq$ has a VNM utility representation U that is determined by any given well-behaved welfare measure W satisfying CIRA and by the constant intrinsic risk proneness ρ_W via*

$$U = (e^{\rho_W W} - 1)/\rho_W.$$

If $\rho_W = 0$, i.e., if $\succeq$ is intrinsic risk neutral, then '$(e^{\rho_W W} - 1)/\rho_W$' stands for W $(= \lim_{\rho \to 0} \log(e^{\rho W} - 1)/\rho)$.

Classical risk proneness $\rho_{AP} = \frac{U''}{U'}$ also has a decomposition, which shows how it combines risk-attitudinal and welfare-related aspects:

Proposition 2 *The classical risk proneness ρ_{AP} of a well-behaved observed order $\succeq$, for $X \subseteq \mathbb{R}$, is determined by any given well-behaved welfare measure W and by the intrinsic risk proneness ρ_W via*

$$\rho_{AP} = \frac{W''}{W'} + W' \rho_W.$$

[8] As $\succeq$ is well-behaved, $\frac{U''}{U'}$ is well-defined, i.e., U exists and is twice differentiable with $U' \neq 0$.

[9] Well-definedness of $\frac{d^2U/dW^2}{dU/dW}$ means that U is twice differentiable in W with nowhere zero first derivative in W, more precisely that U is writable as $\phi(W)$ for a (unique) function $\phi : Rg(W) \to \mathbb{R}$ that is twice differentiable with nowhere zero ϕ' (in which case dU/dW stands for $\phi'(W)$ and d^2U/dW^2 stands for $\phi''(W)$). Well-definedness follows from the well-behavedness of $\succeq$ and W (in fact, ϕ' is everywhere positive, as can be seen via Lemma 1).

Thus the classical risk proneness $\rho_{AP} = \frac{U''}{U'}$ (the growth rate of marginal utility) is the sum of

- a 'welfare component' $\frac{W''}{W'}$, the growth rate of marginal welfare, and
- a 'risk component' $W'\rho_W$, the intrinsic risk proneness weighted by marginal welfare. The weighting by marginal welfare is plausible, as it reflects the intuition that the attitude to welfare-risk (captured by ρ_W) matters only to the extent that welfare varies (captured by W').

Proposition 2 does not require CIRA. Without CIRA, the intrinsic welfare risk ρ_W is not constant and VNM utility U is not given by $(e^{\rho_W W} - 1)/\rho_W$, but U still obeys the differential equation '$\frac{U''}{U'} = \frac{W''}{W'} + W'\rho_W$'. This differential equation cannot be solved analytically (except in special cases such as that of constant ρ_W), but it shows that U is always determined by the two functions W and ρ_W, with or without CIRA. Hence the central conceptual point – that classic utility has two determinants, welfare and intrinsic risk attitude – does not require CIRA.

4 Uniquely revealed welfare and intrinsic risk attitude

So far, the welfare measure W – and hence also the intrinsic risk proneness ρ_W and our decomposition of classic utility and classic risk proneness – are empirically underdetermined: they are theoretic constructs that are only partially revealed by the ordinal information $\succsim$. More precisely, the welfare measure W in Proposition 1 has three remaining degrees of freedom: the intrinsic risk proneness ρ and the two degrees of freedom implicit in the choice of VNM representation U. Surprisingly, full uniqueness can be achieved by adding two simple hypotheses about welfare, namely a range condition and a normalisation condition. We begin with the range condition:

Full-range: There exist arbitrarily good or bad situations. That is, for all welfare levels $w \in \mathbb{R}$ there is a situation $x \in X$ such that $W(x) = w$.

Full-range is a richness assumption on the set of situations considered: this set should include situations of arbitrary quality, be these situations feasible or merely theoretic. Note that VNM utility could still be bounded below or above. Implicitly, Full-range is also a condition on the scale on which welfare is measured: that scale should include all real numbers as meaningful welfare levels. We return to measurement-theoretic questions later.

To normalise the welfare measure, we consider a fixed reference situation $\bar{x} \in X$, representing for instance a 'poverty point'. We call the welfare function W, or any other function from X to $\mathbb{R}$, *normalised* if at the reference point $\bar{x}$ it takes the value 0 and has a derivative of size 1. The derivative of W, or *marginal welfare*, captures how small changes of the situation affect welfare.[10]

Normalisation: The welfare function W is normalised.

[10] In the basic $X \subseteq \mathbb{R}$, the size W' is the absolute value $|W'|$, which normally equals W' as $W' > 0$, i.e., as 'more is better'. In the general case $X \subseteq \mathbb{R}^k$ $(k \geq 1)$, the size of $W' = \left(\frac{dW}{dx_1}, \ldots, \frac{dW}{dx_k}\right)$ is the length $\|W'\|$.

Normalisation requires measuring welfare on a scale that sets welfare to 0 and marginal welfare to size 1 at the reference point. Measurement scales are conventions, not substantive assumptions. The scale fixes the meaning of numbers, i.e., informally, the mapping from numbers to meanings of numbers. One can always scale welfare in accordance with Normalisation: any well-behaved welfare function satisfying CIRA and Full-range can be transformed into one satisfying also Normalisation, by applying an increasing affine transformation. Rescaling a welfare function changes absolute welfare levels and welfare differences. This scale-dependence does not make levels and differences meaningless: statements such as 'welfare is 2' and 'welfare rises by 3' have substantive meanings, but meanings that are *relative* to the chosen scale. The axiom of Normalisation becomes less innocent when one engages in interpersonal comparisons of welfare levels and/or differences. This point will be discussed later, when we generalise Normalisation.

We now state our central theorem. It states that there exists a unique, and therefore revealed, welfare measure satisfying our conditions. It is obtained by choosing U and ρ in particular ways in the formula of Proposition 1. The result will assume that the observed order $\succeq$ is *broad-ranging*. By this we mean that for any situations there exist much better or much worse situations, more precisely for any situations $x, y \in X$ with $x \succeq y$ there exists a situation $z \in X$ such that $z_{\frac{1}{2}}y_{\frac{1}{2}} \succ x$ or $y \succ z_{\frac{1}{2}}x_{\frac{1}{2}}$. Note that '$z_{\frac{1}{2}}y_{\frac{1}{2}} \succ x$ or $y \succ z_{\frac{1}{2}}x_{\frac{1}{2}}$' means that z is so good that its 50-50 mixture with y beats x or so bad that its 50-50 mixture with x loses to y. This condition holds under most models of preferences under risk, for instances under all HARA preferences.

Definition 3 *Given an observed order $\succeq$, if there is a unique well-behaved welfare measure satisfying CIRA, Full-range and Normalisation, then it is called* revealed *by $\succeq$ and denoted $W_{\succeq}$, and the intrinsic risk proneness w.r.t. $W_{\succeq}$ is called* revealed *by $\succeq$ and denoted $\rho_{\succeq}$.*

Theorem 1 *Any well-behaved and broad-ranging observed order $\succeq$ reveals a welfare measure $W_{\succeq}$ and a constant intrinsic risk proneness $\rho_{\succeq}$, given by*

$$W_{\succeq} = \log(\rho_{\succeq}U + 1)/\rho_{\succeq}$$

and

$$\rho_{\succeq} = \begin{cases} \frac{-1}{\sup U} \ (<0) & \text{if } \sup U \neq \infty & \text{(intrinsic risk aversion)} \\ \frac{-1}{\inf U} \ (>0) & \text{if } \inf \neq -\infty & \text{(intrinsic risk proneness)} \\ 0 & \text{if } \sup U = \infty \text{ and } \inf U = -\infty & \text{(intrinsic risk neutrality)} \end{cases}$$

where U is the (unbounded) normalised VNM representation of $\succeq$.

The utility function U could for instance be one of the many HARA utility functions, which are used in applications and are indeed unbounded.

Like the classical measure of risk attitude ρ_{AP}, our measure $\rho_{\succeq}$ can be calculated from the VNM representation U, but the calculation looks very different: while ρ_{AP} $(= \frac{U''}{U'})$ is derived from the curvature of U, $\rho_{\succeq}$ is derived from extreme values of U.

5 Application, discussion and generalisation

In this final section, we first show how our welfare measure can be combined with empirical data to yield practical welfare assessments. We then turn to *social* welfare – the starting point the Harsanyi-Sen debate. This will lead to a closer analysis of our hypotheses, and to a generalisation of these hypotheses and the theorem.

Empirical application

It is natural to combine our expression for welfare with empirically supported VNM utility functions. For example, let $X = (0, \infty)$ and assume the agent is found empirically to display decreasing absolute risk aversion, with a constant *relative* risk aversion ('CRRA') of $\eta \geq 0$. Many studies confirm this picture, although the value of η is highly context-dependent. The agent's normalised utility function is then of the well-known CRRA type:

$$U(x) = \frac{\bar{x}}{1-\eta}\left(\left(\frac{x}{\bar{x}}\right)^{1-\eta} - 1\right) \text{ for all } x \in X.$$

If $\eta = 1$, this formula has the usual interpretation, as $U(x) = \bar{x}\log\frac{x}{\bar{x}}$ $(= \lim_{\eta\to 1}\frac{\bar{x}}{1-\eta}\left(\left(\frac{x}{\bar{x}}\right)^{1-\eta} - 1\right))$. By Theorem 1, the agent has revealed intrinsic risk proneness given by[11]

$$\rho_{\succeq} = \frac{1-\eta}{\bar{x}}$$

and revealed welfare given by[12]

$$W_{\succeq}(x) = \bar{x}\log\frac{x}{\bar{x}} \text{ for all } x \in X.$$

Thus, the CRRA model leads to welfare of a simple logarithmic form, which does not depend on any parameter, except the reference situation $\bar{x}$ used for the normalisation. Interestingly, the debate about the value of relative risk aversion η does not affect the measurement of welfare – it only affects the intrinsic risk attitude $\rho_{\succeq} = \frac{1-\eta}{\bar{x}}$.

Other models than the CRRA model – for instance other HARA models – lead to other concrete formulas for welfare and intrinsic risk proneness via Theorem 1. The simple computations are left to readers.

Social welfare

Part of why measuring individual welfare matters is that it allows us to measure *social* welfare, as a guide for policy making. Let us focus here on *utilitarian* social welfare: social welfare is sum-total individual welfare. If John Harsanyi and followers are right, then individual welfare is simply VNM utility, and one should maximise sum-total VNM utility. If the critics such as Amartya Sen and John Weymark are right, then individual welfare differs from VNM utility, and one should *not* maximise

[11] Check this by distinguishing between the cases $\eta > 1$ (where $\sup U < \infty$), $\eta < 1$ (where $\inf U > -\infty$) and $\eta = 1$ (where $\sup U = \infty$ and $\inf U = -\infty$).

[12] Because $W(x) = \frac{\log(\rho U(x)+1)}{\rho} = \frac{\log\left(\rho\frac{1}{\rho}\left(\left(\frac{x}{\bar{x}}\right)^{\rho\bar{x}}-1\right)+1\right)}{\rho} = \frac{\log\left(\left(\frac{x}{\bar{x}}\right)^{\rho\bar{x}}\right)}{\rho} = \bar{x}\log\frac{x}{\bar{x}}$.

total utility, but total welfare. We adopt the latter view, and operationalise welfare via Theorem 1.

Concretely, consider a society of individuals $i = 1, ..., n$ ($n \geq 2$). Each individual i satisfies the (observable) assumptions of Theorem 1, and so we can measure her welfare by $W_i = \log(\rho_i U_i + 1)/\rho_i$, where U_i is her (normalised) VNM utility function and ρ_i is her revealed intrinsic risk proneness.

What goes wrong when maximising total utility $\sum_i U_i$ (or total weighted utility[13]) rather than total welfare $\sum_i W_i$? Under the plausible assumption that individuals are intrinsic risk averse, each W_i is a convex transformation of U_i by Theorem 1, and so each U_i is a concave transformation of W_i. Thus, maximising total utility (or weighted utility) means maximising the total of *concavely transformed* welfare levels – an approach that effectively prioritises the worse-off and is therefore known as *prioritarianism* and regarded as a major alternative to utilitarianism [citations TBA]. Ironically, this implies that John Harsanyi, officially a dedicated utilitarian, is effectively a prioritarian, and that his 'utilitarian' theorem supports prioritarianism.[14]

How should the utilitarian principle of maximising total welfare handle risky prospects in $\mathcal{P}$ rather than riskless situations in X? This is notoriously controversial. Ex-post utilitarians maximise total expected ex-post welfare $\sum_i \mathbb{E}_p W_i$. Ex-ante utilitarians instead maximise total ex-ante welfare $\sum_i \mathcal{W}_i(p)$, where $\mathcal{W}_i$ is the extension of i's welfare measure W_i to $\mathcal{P}$ such that a risky prospect gives the same welfare as a certainty equivalent. Either version of utilitarianism respects exactly one of the two requirements in Harsanyi's 'utilitarian' theorem: ex-post utilitarianism respects social VNM rationality, while ex-ante utilitarianism respects Pareto. Utilitarians thus face a hard choice when defining their principle under risk. Whichever approach is taken, it can be operationalised using our welfare measure.

A generalisation

We now discuss potential criticisms of the three hypotheses, which will lead us to generalise these hypotheses and the theorem. We begin with Full-range, followed by Normalisation and CIRA.

Full-range discussed and generalised

By requiring that W ranges over the full set $\mathbb{R}$, this hypothesis assumes not only that there are arbitrarily good or bad situations in X, but also that all real numbers are *meaningful* as welfare levels. The latter is a condition on the choice of measurement scale: that scale must has range $\mathbb{R}$. But one might want to measure welfare on a different scale. For instance, if a scale with range $\mathbb{R}$ is transformed exponentionally, by replacing ('relabelling') any welfare level $w \in \mathbb{R}$ with $e^w \in (0, \infty)$, then the new scale has range $(0, \infty)$. To allow for such other measurement scales, we now introduce an arbitrary (non-empty and open) interval $D \subseteq \mathbb{R}$ of *meaningful* welfare levels, e.g., $D = \mathbb{R}$ or $D = (0, \infty)$ or $D = (0, 1)$, and impose the following hypothesis (which reduces to Full-range if $D = \mathbb{R}$):

[13] Total weighted utility is $\sum_i \alpha_i U_i$ for some weights $\alpha_i > 0$.

[14] By his theorem, a Pareto condition and social VNM rationality implies maximising the sum of (suitably scaled) individual VNM utility functions.

Full-range$_D$: There are situations of arbitrary quality in D, i.e., $\{W(x) : x \in X\} = D$.

Normalisation discussed and generalised

This condition sets welfare to 0 and marginal welfare size to 1, at a given reference point $\bar{x}$, for instance a poverty level. More generally, one might fix numbers $r \in D$ and $s > 0$, and impose the following hypothesis (which reduces to Normalisation if $r = 0$ and $s = 1$):

Normalisation$_{r,s}$: At the reference point $\bar{x}$, the welfare is r and the marginal welfare size is s, i.e., W takes value r and has a derivative of size s.

What speaks for replacing Normalisation with Normalisation$_{r,s}$? Normalisation is questionable when one makes interpersonal comparisons of welfare, since Normalisation treats everyone as having the same welfare (and marginal welfare size) at $\bar{x}$. Nothing is wrong with assuming Normalisation for for a *given* person – this just means one must measure welfare on a scale such that (for instance) the welfare denoted by '0' coincides with that experienced by the person at $\bar{x}$. But assuming Normalisation for many persons simultaneously leads to questionable welfare comparisons at $\bar{x}$. By contrast, Normalisation$_{r,s}$ works even in an interpersonal context, because r and s can be person-dependent.

In some contexts, Normalisation *is* defensible. Why? First, recall that the status and choice of normalisation is an old problem in welfare economics, although it is usually raised for VNM utility rather than our welfare measures. The sensitivity of interpersonal utility comparisons and total-utility-maximisation to the normalisation of individual utilities has already bothered Harsanyi, and has led to different concrete proposals. Some favour a '0-1 normalisation' that sets utility to 0 resp. 1 at two reference outcomes (e.g., Isbell 1959, Segal 2000, Adler 2012, 2016). Others favour a normalisation with a single reference outcome at which both utility and marginal utility are fixed, say to 0 resp. 1 (e.g., Fleurbaey and Zuber 2021). Our hypothesis Normalisation corresponds to the latter proposal, modulo the difference between VNM utility and welfare.

In general, there are at least three contexts in which Normalisation seems appropriate:

1. Affinely measurable welfare: Suppose we pursue a lower ambition by aiming to measure welfare on an affine rather than absolute scale. The informational content of W is then more limited: it lies between absolute and purely ordinal information. Values and differences of W are no longer significant, but ratios of differences are significant (just as for VNM utility[15]). Normalisation is then unproblematic, since every welfare function is affinely equivalent to one satisfying Normalisation. For affine welfare, we lose the possibility of interpersonal comparisons of levels or differences,

[15] VNM utility is also affine, but it measures something else than welfare.. So, ratios of differences of VNM utility are also unique, but they are hard to interpret. While '$\frac{W(x)-W(y)}{W(x')-W(y')} = 1$' means that the change from x to y gives as much welfare as that from x' to y', '$\frac{U(x)-U(y)}{U(x')-U(y')} = 1$' does not say something about welfare alone but about the interplay of welfare and risk attitude.

and of the mentioned utilitarian social welfare function '$\sum_i W_i$', which requires comparisons of differences. But other welfare function remain possible, most notably the Nash social welfare function.[16]

2. Contextualised welfare: Suppose the question is not what welfare individuals have intrinsically, but what welfare they should be *treated* as having in a given context of interpersonal comparisons, welfare aggregation, commodity allocation or other policy choice. For such a 'contextualised' welfare, Normalisation says that we should *treat* individuals as having the same welfare (of 0) and the same marginal welfare size (of 1) at a given reference point $\bar{x}$. But why? Normalisation ensures that Pigou-Dalton transfers increase social welfare $\sum_i W_i$. More precisely, if each W_i is increasing and concave on $X_i \subseteq \mathbb{R}$, and satisfies Normalisation where the reference point is a common 'poverty point' $\pi \in \mathbb{R}$ below/above which someone counts as poor/rich, then social welfare $\sum_i W_i$ increases by transferring resources from rich to poor persons.[17] So, Normalisation gives utilitarianism an unexpected egalitarian appeal. This egalitarian argument for the normalisation Normalisation is introduced and developed axiomatically in Fleurbaey and Zuber (2021), in a different version with VNM utility instead of welfare.[18] In their words, Normalisation leads to 'fair utilitarianism'.

3. Locally objective welfare: One often distinguishes between 'objective' and 'subjective' notions of welfare [citations TBA]. Objective welfare is given by 'objective' features such as wealth or consumption levels, and subjective welfare by 'subjective' features such as tastes or experienced happiness. Technically, someone's objective welfare depends only on the situation in X, while subjective welfare depends also on her tastes or other subjective features. Of course, a notion of welfare can of hybrid subjective-objective type, and the extent of objectivity can moreover vary between different situations in X. Our analysis is open to objective as well as subjective approaches. Now Normalisation can be regarded as reflecting a *local objectivity* of welfare: the welfare level becomes objective at the reference point $\bar{x}$. For instance, a situation of high misery $\bar{x}$ might give an *objective* (low) amount of welfare and (high) amount of marginal welfare. This would justify Normalisation. But full-blown objectivity at $\bar{x}$ is not needed for Normalisation. Welfare could be of subjective type and merely become 'effectively objective' at $\bar{x}$ in the sense that the relevant subjective features become the same for everyone at $\bar{x}$. So, welfare at $\bar{x}$ is determined by subjective but universal features. For instance, a situation of high misery $\bar{x}$ could be disliked by everyone to the same (large) extent, because (so the argument) differences in taste only arise once the most basic physical needs are covered. So to say, the external circumstances take over at $\bar{x}$ by crowding out the subjective differences. The plausibility of such local (effective) objectivity of welfare is certainly debatable. The plausibility depends partly on what counts as a 'situation'. Mere wealth levels are perhaps too uninformative 'situations' for local objectivity to emerge at a 'poverty level' or other

[16] The latter is definable as $\prod_i (W_i - W_i(\bar{x}))^{1/n}$, and is restricted to situations in $\{x \in X : W_i(x) \geq 0$ for all $i\}$. Bossert and Weymark (2004) review various social welfare functions and their underlying informational requirements.

[17] That is, for all situations $x, y \in X$ and individuals j, k, if $x_j < y_j < \pi < y_k < x_k$, $y_j - x_j = x_k - y_k$, and $x_l = y_l$ for all other individuals k, then $\sum_i W_i(x) < \sum_i W_i(y)$.

[18] While in their context concavity represents classical risk-aversion, in our context it represents decreasing marginal welfare. In both cases, concavity is plausible.

reference point. This might change if situations are detailed consumption vectors, or even entire 'lives' or Sen-type functioning vectors. The more information is packed into situations, perhaps including 'quasi-subjective' information, the less room is left for subjectivity in welfare assessments. Stigler and Becker's thesis 'De gustibus non est disputandis' and Sen's programme of evaluating fine-grained functioning vectors can be regarded as two (very different) attempts to approach objectivity of evaluation through a suitable level of description of the objects of evaluation (Stigler and Becker 1977, Sen 1985).

CIRA discussed and generalised

CIRA requires a constant aversion to *intrinsic* risk: if all possible *welfare* outcomes of a risky prospect are shifted by the same amount, then the prospect's equivalent welfare is also shifted by this amount. We have suggested that CIRA is a plausible explanation for why classic Arrow-Pratt absolute risk aversion is often found to be decreasing: this empirical finding follows from CIRA, assuming that marginal welfare is diminishing.

Here is a possible defence of CIRA. For concreteness, let situations be wealth levels, where $X \subseteq \mathbb{R}$. The agent has some initial wealth. Wealth prospects describe final-wealth probabilities. Assume that, firstly, the agent is dynamically consistent in that her ranking of final-wealth prospects would not change if her initial wealth changed first; and ,secondly, the agent always perceives and ranks wealth prospects in terms of gains or losses of welfare rather than final levels of welfare. The focus on gains or losses rather than levels follows Kahnemann-Tversky's prospect theory, but in an (arguably more plausible) version based on welfare rather than monetary wealth. These two assumptions can be shown to imply CIRA.[19]

Be this as it may, if one finds CIRA too restrictive, one can replace it with a more flexible hypothesis that can accommodate CIRA as well as various other interesting welfare-based hypotheses, such as constant *relative* intrinsic risk, i.e., constant risk in *welfare ratios*. While CIRA requires a constant aversion to risk *in welfare simpliciter*, the generalisation requires a constant aversion to risk in some welfare-based quantity,

[19] More precisely, let $\succeq_z$ be the order on $\mathcal{P}$ that the agent would hold if her initial wealth were z ($\in X$); it equals $\succeq$ if z is her true initial wealth. *Assumption 1* ('status quo independence' or 'dynamic consistency'): $\succeq_z$ does not depend on z. Now, given an initial wealth level z, each final-wealth lottery $p \in \mathcal{P}$ yields a welfare-*gain* lottery $\gamma_{p|z}$ over $\mathbb{R}$, where of course the probability of a welfare gain of $g \in \mathbb{R}$ equals the probability that final welfare exceeds initial welfare $W(z)$ by g, i.e., $\gamma_{p|z}(g) = p(W(.) - W(z) = g)$. *Assumption 2* ('welfare-gain-based preferences'): there is an order $\succeq^*$ over welfare-gain lotteries (i.e., lotteries on $\mathbb{R}$ with finite support) such that $p \succeq_z q \Leftrightarrow \gamma_{p|z} \succeq^* \gamma_{q|z}$ for all final-wealth lotteries $q, q \in \mathcal{P}$ and initial wealth levels $z \in X$. *Claim*: Assumptions 1 and 2 imply CIRA, supposing well-behavedness of $\succeq$ and W and Full-range. *Proof sketch*: Let Δ, p, w_p, q, w_q be as in CIRA. Let $p(W = w) = q(W = w + \Delta)$ for all $w \in \mathbb{R}$. We show $w_q = w_p + \Delta$. Let z_0 be the true initial wealth, and z the wealth level such that $W(z) = W(z_0) + \Delta$ (z exists by Full-range). Let $x_p, x_q \in X$ be certainty equivalents of p resp. q (they exist as $\succeq$ is well-behaved). Now $p \sim x_p$, i.e., $p \sim_{z_0} x_p$; so by Ass. 2 $\gamma_{p|z_0} \sim^* W(x_p) - W(z_0) = w_p - W(x_0)$ (identifying the welfare gain $W(x_p) - W(z_0) = w_p - W(x_0)$ with a riskless welfare-gain lottery). Further, $q \sim x_q$, hence by Ass. 1 $q \sim_z y$; so by Ass. 2 $\gamma_{q|z} \sim^* W(x_q) - W(z) = w_q - W(z)$. The welfare-gain lotteries $\gamma_{p|z_0}$ and $\gamma_{q|z}$ coincide, informally because q's welfare prospect equals p's shifted by Δ and z's welfare equals z_0's shifted by Δ. Since $\gamma_{p|z_0} \sim^* w_p - W(z_0)$, $\gamma_{q|z} \sim^* w_q - W(z)$, and $\gamma_{p|z_0} = \gamma_{q|z}$, we have $w_p - W(z_0) \sim^* w_q - W(z)$. It follows that $w_p - W(z_0) = w_q - W(z)$, because only identical welfare gains are indifferent (as $\succeq$ and W are well-behaved). Thus $w_q - w_p = W(z) - W(z_0) = \Delta$. ■

i.e., some *transformation of welfare*. Formally, we replace CIRA with the following condition that is defined relative to a welfare transformation τ, which can be any smooth function τ from the mentioned interval D of meaningful welfare levels onto $\mathbb{R}$ with $\tau' > 0$:

CIRA$_\tau$: If all transformed welfare outcomes of a prospect rise by the same amount, then the transformed equivalent welfare also rises by this amount. Formally, $Rg(W) \subseteq D$ and, for all $\Delta > 0$ and all prospects $p, q \in \mathcal{P}$ with equivalent welfare w_p resp. w_q, if $p(\tau(W) = t) = q(\tau(W) = t + \Delta)$ for all $t \in \mathbb{R}$ then $\tau(w_q) = \tau(w_p) + \Delta$.

CIRA$_\tau$ reduces to CIRA if $\tau(w) = w$ for all w in $D = \mathbb{R}$. If instead $\tau(w) = \log w$ for all w in $D = (0, \infty)$, then CIRA$_\tau$ requires a constant aversion to risk in welfare *ratios*, i.e., a constant *relative* intrinsic risk aversion.

The theorem generalised

Even in their generalised form, our hypotheses lead to a unique welfare measure:

Theorem 2 *Given a well-behaved and broad-ranging observed order $\succeq$, there exists a unique well-behaved welfare measure $W : X \to \mathbb{R}$ satisfying CIRA$_\tau$, Full-range$_D$ and Normalisation$_{r,s}$ (for given parameters τ, D, r, s), given by*

$$W = \tau^{-1}(\log(\rho s \tau'(r) U + 1)/\rho + \tau(r)),$$

where $U : X \to \mathbb{R}$ is the (unbounded) normalised VNM representation of $\succeq$ and

$$\rho = \begin{cases} \frac{-1}{s\tau'(r) \sup U} < 0 & \text{if } \sup U \neq \infty \\ \frac{-1}{s\tau'(r) \inf U} > 0 & \text{if } \inf \neq -\infty \\ 0 & \text{if } \sup U = \infty \text{ and } \inf U = -\infty. \end{cases}$$

Theorem 1 is a special case of Theorem 2, obtained if $D = \mathbb{R}$, $r = 0$, $s = 1$, and τ is the identity transformation, because the three hypotheses then reduce to the original ones and the formula for W reduces to that in Theorem 1.[20]

As an application, assume $D = (0, \infty)$ and $\tau = \log$, so that CIRA$_\tau$ requires constant *relative* intrinsic risk aversion. Then the formula for welfare reduces to $W = r(\frac{\rho s}{r} U + 1)^{1/\rho}$. So, welfare is now a geometric rather than logarithmic function of VNM utility.

6 Conclusion

We have shown that plausible working hypotheses allow one to operationalise the difficult notion of individual welfare, and to disentangle an agent's ordinary VNM utility into two determinants, namely her welfare or intrinsic utility and her intrinsic risk attitude. This makes welfare and intrinsic risk attitude indirectly observable quantities, and suggests an explanation of the empirical phenomenon of decreasing absolute risk aversion in terms of an interplay of diminishing marginal *welfare* and constant

[20] If $\rho = 0$ then the expression $\log(\rho s \tau'(r) U + 1)/\rho$ in the formula should be read as $s\tau'(r)U$ $(= \lim_{\rho \to 0} \log(\rho s \tau'(r) U + 1)/\rho)$.

intrinsic risk aversion. While we have suggested reasons for adopting our hypotheses as working assumptions, the hypotheses remain debatable. We have presented generalised hypotheses, leading to a generalised formula for welfare.

Social welfare analysis can now use a more satisfactory observable measure of individual welfare than VNM utility. Social decisions often require not only adequate measures of individual and social welfare in fixed situations, but also an adequate handling of risk. Handling risk is far from obvious, as is highlighted by the difference between ex-ante and ex-post utilitarianism. In essence, one needs a *social* attitude to *intrinsic* risk. One approach is to aggregate not only individual into social welfare, but also individual into social intrinsic risk attitudes. Since both individual characteristics – welfare and intrinsic risk attitude – are observably 'contained' in individual preferences under risk, this approach could be pursued within the standard framework of preference aggregation under risk.

Appendix

The generalised setup with arbitrary alternatives

The main text took the set of alternatives X to be a (non-empty, open connected) subset of $\mathbb{R}^k$ for some $k \geq 1$. All hypotheses and results continue to hold as such for an arbitrary non-empty set X. This requires generalising the notions of 'regular' and 'normalised' functions on X – the two notions that refer to derivatives, which do not exist in general.

Regularity generalised. For a function $W : X \to \mathbb{R}$, the main text took 'regular' to mean that W is smooth and has a nowhere zero derivative. In general, the set of *regular* functions is defined as any given set $\mathcal{F}$ of functions $f : X \to \mathbb{R}$ such that, for all $f \in \mathcal{F}$, the range $Rg(f) = \{f(x) : x \in X\}$ is an open interval and, for every strictly increasing $\phi : Rg(f) \to \mathbb{R}$, we have $\phi \circ f \in \mathcal{F}$ if and only if ϕ is smooth with $\phi' > 0$. This condition is met by the main text's notion of 'regular', as shown below.

Normalisation generalised. In the main text, a function $W : X \to \mathbb{R}$ counts as normalised if has value 0 and a derivative of size 1 at the reference point $\overline{x}$. As derivatives are undefined in general, we now define a generalised notion of normalised functions. It is given by a set $\mathcal{N}$ of functions $f : X \to \mathbb{R}$, called the *normalised* functions, satisfying minimal conditions: (i) for all $f \in \mathcal{N}$, $f(\bar{x}) = 0$; (ii) for all $f \in \mathcal{N}$ and all smooth transformations $\phi : Rg(f) \to \mathbb{R}$ with $\phi(0) = 0$, we have $\phi \circ f \in \mathcal{N} \Leftrightarrow \phi'(0) = 1$; (iii) each regular function $f \in \mathcal{F}$ is normalisable, i.e., has an increasing affine transformation in $\mathcal{N}$. Condition (ii) is our abstract analogue of the condition that normalised functions have derivative size 1 and $\overline{x}$. Intuitively, normalised functions have the right value at $\overline{x}$ by (i), and have the same (abstract) derivative size at $\overline{x}$ by (ii).[21]

In general, a function $f : X \to \mathbb{R}$ is *normalisable* if it has an increasing affine transformation g in $\mathcal{N}$, in which case g is the *normalisation* of f. Normalisations

[21] In (ii), $\phi \circ f$ intuitively has the same abstract derivative at $\overline{x}$ as f if and only if $\phi'(0) = 1$. Reason: $(\phi \circ f)'(\overline{x}) = \phi'(f(\overline{x}))f'(\overline{x}) = \phi'(0)f'(\overline{x})$, assuming abstract derivatives behave like ordinary ones, and (i) holds.

exist at least for regular functions by (iii), and are always unique.[22]

The axiom of Normalisation ('W is normalised') now makes sense in general. To also make general sense of the axiom of Normalisation$_{r,s}$ ('W has value r and a derivative of size s at $\overline{x}$'), we must generalise the meaning of 'size of derivative'. While $\mathcal{N}$ does not induce a notion of derivative, it does induce a notion of derivative *size*. How? Recall that all functions in $\mathcal{N}$ intuitively have derivative size 1 at $\overline{x}$. We define the *(abstract) derivative size* of any normalisable function $f : X \to \mathbb{R}$ as the number s in the representation $f = sg + r$ where $g \in \mathcal{N}$ is the normalisation of f and where $s > 0$ and $r \in \mathbb{R}$.[23] This definition ensures that functions in $\mathcal{N}$ (like g) have derivative size 1, and that the derivative size of $f = sg + r$ equals s times that of g.

Our generalised setup includes the concrete setup of the main text as a special case:

Lemma 1 *If, as in the main text, X is a non-empty open connected subset of $\mathbb{R}^k$ ($k \geq 1$) and the sets of regular and normalised functions are, respectively,*

$$\mathcal{F} = \{f : X \to \mathbb{R} : f \text{ is smooth, } f'(x) \neq \mathbf{0} \text{ for all } x \in X\}$$
$$\mathcal{N} = \{f : X \to \mathbb{R} : f(\overline{x}) = 0,\ f'(\overline{x}) \text{ exists and is of size 1}\}$$

then the above conditions on $\mathcal{F}$ and on $\mathcal{N}$ are satisfied.

Proof. Let X, $\mathcal{F}$ and $\mathcal{N}$ be as in the main text. We first establish the conditions that $\mathcal{F}$ (part 1), and then those for $\mathcal{N}$ (part 2).

1. Fix an $f \in \mathcal{F}$. The range $Rg(f)$ is an interval because continuous images of connected sets are connected. This interval is open, as one easily deduces from the fact that f has a non-zero derivative at all $x \in X$. Now fix a strictly increasing $\phi : Rg(f) \to \mathbb{R}$. By basic calculus, if ϕ is smooth with $\phi' > 0$, then $\phi \circ f \in \mathcal{F}$.

Henceforth we assume $\phi \circ f \in \mathcal{F}$ and must show that ϕ is smooth with $\phi' > 0$. Put $g = \phi \circ f$.

Claim 1: If $X \subseteq \mathbb{R}$ (i.e., $k = 1$), then f^{-1} exists and is smooth.

Let $X \subseteq \mathbb{R}$. As $f \in \mathcal{F}$, the derivative f' exists and is continuous and nowhere zero. So f' is everywhere positive or everywhere negative. Thus f is strictly monotonic, hence invertible. To show that $h = f^{-1}$ is smooth, we show by induction that for all $n \geq 1$ the n^{th} derivative $h^{(n)}$ exists and is a ratio $\frac{a}{b}$ of smooth functions $a, b : X \to \mathbb{R}$ with $b > 0$. First consider $n = 1$. As $f' > 0$, the function, $h' = (f^{-1})'$ exists and equals $\frac{1}{f'(h)}$, a ratio of the claimed form. Now let $n > 1$ and assume that $h^{(n-1)}$ exists and is a ratio of the claimed form, say $h^{(n-1)} = \frac{a}{b}$. By implication, $h^{(n)}$ exists and equals $\frac{a'b-b'a}{b^2}$, which is again a ratio of the claimed form. Qed.

Claim 2: If $X \subseteq \mathbb{R}$ (i.e., $k = 1$), then ϕ is smooth with $\phi' > 0$.

[22] *Proof of uniqueness*: Assume $f, af + b \in \mathcal{N}$, where $a > 0$ and $b \in \mathbb{R}$. We must show that $a = 1$ and $b = 0$. By (i), $b = 0$. Applying (ii) with ϕ given by $t \mapsto at$, we have $a = 1$, since $\phi'(0) = a$ and $\phi \circ f \in \mathcal{N}$.

[23] In this representation of f, all of g, s and r are unique: g is the (unique) normalisation of f, r is given by $r = f(\overline{x})$, and s is given by $s = (f(x) - r)/g(x)$ for any $x \in X$ that is chosen such that $g(x) \neq 0$ (an x with $g(x) \neq 0$ exists because otherwise g would be the zero function, although this function is not in $\mathcal{N}$ by the condition (ii) applied with ϕ taking always the value 0).

Assume $X \subseteq \mathbb{R}$. As $g = \phi \circ f$ and f is (by Claim 1) invertible, we have $\phi = g \circ h$, where $h = f^{-1}$. The smoothness of ϕ can be deduced from the fact that $\phi = g \circ h$ and that g and (by Claim 1) h are smooth. We skip the complete inductive argument. In short, ϕ' exists and equals $h'g'(h)$; so ϕ'' exists and equals $h''g'(h) + h'(g'(h))' = h''g'(h) + h'^2 g''(h)$; and so on for higher derivatives of ϕ.

To see why $\phi' > 0$, fix a $w \in Rg(f)$. Pick an $x \in X$ such that $f(x) = w$. We have $g'(x) = \phi'(w)f'(x)$ since $g'(x) = (\phi \circ f)'(x) = \phi'(f(x))f'(x) = \phi'(w)f'(x)$. So, as $g'(x)$ and $f'(x)$ are non-zero and (by ordinal equivalence of f and g) of same sign, we have $\phi'(w) > 0$. Qed.

Claim 3: ϕ is smooth with $\phi' > 0$ (completing the proof).

Now we allow X to be multi-dimensional: X is any connected, open, and non-empty subset of $\mathbb{R}^k$ where $k \geq 1$. Let $t \in Rg(f)$. We must show that, at t, ϕ is smooth with $\phi' > 0$. Pick an $x \in f^{-1}(t)$. Since $f'(x) \neq \mathbf{0}$, we may pick a coordinate $j \in \{1, ..., k\}$ such that $\frac{df}{dx_j}(x) \neq 0$. As f is smooth and f' is nowhere zero, there is an open interval $\tilde{X}$ containing x_j such that, for all $y \in \tilde{X}$, $(x_1, ..., x_{j-1}, y, x_{j+1}, ..., x_k) \in X$ and $\frac{df}{dx_j}(x_1, ..., x_{j-1}, y, x_{j+1}, ..., x_k) \neq 0$. Consider f as a function of the j^{th} coordinate in $\tilde{X}$. That is, define the function $\tilde{f} : \tilde{X} \to \mathbb{R}$ given by $y \mapsto \tilde{f}(y) = f(x_1, ..., x_{j-1}, y, x_{j+1}, ..., x_n)$. Let $\tilde{\phi}$ be the restriction of ϕ to $Rg(\tilde{f})$ $(\subseteq Rg(f))$. We now replace the primitives X, f, ϕ and $\mathcal{F}$ with, respectively, $\tilde{X}$, $\tilde{f}$, $\tilde{\phi}$ and $\tilde{\mathcal{F}} = \{s : \tilde{X} \to \mathbb{R} : s \text{ is smooth } \& \ s'(x) \neq 0 \text{ for all } x \in \tilde{X}\}$. Note that we indeed have $\tilde{f} \in \tilde{\mathcal{F}}$ (shown using that $f \in \mathcal{F}$) and $\tilde{\phi} \circ \tilde{f} \in \tilde{\mathcal{F}}$ (shown using that $\phi \circ f \in \mathcal{F}$). As $\tilde{X}$ is one-dimensional, Claim 2 applies to these modified primitives. So, $\tilde{\phi}$ is smooth with $\tilde{\phi}' > 0$. Thus, as ϕ coincides with $\tilde{\phi}$ on $Rg(\tilde{f})$, ϕ is smooth with $\phi' > 0$ on $Rg(\tilde{f})$, and hence in particular at t.

2. We show all three conditions on $\mathcal{N}$.

- Condition (i) holds by definition of $\mathcal{N}$.
- To show (iii), fix an $f \in \mathcal{N}$ and a smooth $\phi : Rg(f) \to \mathbb{R}$ with $\phi(0) = 0$. If $\phi'(0) = 1$, then $\phi \circ f \in \mathcal{N}$, because $\phi \circ f(\overline{x}) = \phi(0) = 0$, and because $(\phi \circ f)'(\overline{x})$ exists (as $f'(\overline{x})$ and ϕ' exist) and satisfies $\|(\phi \circ f)'(\overline{x})\| = \|\phi'(f(\overline{x}))f'(\overline{x})\| = |\phi'(f(\overline{x}))| \, \|f'(\overline{x})\| = 1 \times 1 = 1$. If instead $\phi'(0) \neq 1$, then $\phi \circ f \notin \mathcal{N}$, because $\|(\phi \circ f)'(\overline{x})\| \neq 1$.
- To show (iii), fix an $f \in \mathcal{F}$. The increasing affine transformation $g = \frac{1}{\|f'(\overline{x})\|}(f - f(\overline{x}))$ belongs to $\mathcal{N}$, since $g(\overline{x}) = 0$, and $g'(\overline{x})$ exists (as $g'(\overline{x})$ exists) with $\|g'(\overline{x})\| = \frac{1}{\|f'(\overline{x})\|} \|f'(\overline{x})\| = 1$. ■

Proof of Proposition 1

All subsequent proofs will be stated such that they can be read either with the generalised setup in mind or with the main text's concrete setup in mind, depending on the reader's taste. Given a welfare measure W, let $\mathcal{P}^W$ be the set of *welfare prospects*, i.e., finite-support lotteries over $Rg(W)$ rather than X. To each prospect $p \in \mathcal{P}$ corresponds a welfare prospect in $\mathcal{P}^W$, denoted by p^W, where for each $w \in W$ we define $p^W(w)$ as $p(W = w)$, the probability that p results in an outcome with welfare w.

We shall use the classic concept of risk aversion, briefly defined in Section 2.

Lemma 2 *Assume $\succeq$ has a VNM representation $U : X \to \mathbb{R}$ and $W : X \to \mathbb{R}$ is compatible with riskless comparisons. Then:*

(a) For all prospects $p, q \in \mathcal{P}$, $p^W = q^W \Rightarrow p \sim q$.

(b) In particular, we can define an order $\succeq^W$ on $\mathcal{P}^W$ by letting $a \succeq^W b$ if and only if $p \succeq q$ for some (hence by (a) any) $p, q \in \mathcal{P}$ with $p^W = a$ and $q^W = b$.

(c) $\succeq^W$ has a VNM representation, namely the (unique and strictly increasing) function $\phi : Rg(W) \to \mathbb{R}$ such that $U = \phi \circ W$.

(d) $\succeq^W$ displays constant classical risk aversion if and only if CIRA holds.

(e) In particular, if CIRA holds, then ϕ is linear or strictly concave or strictly convex.

Proof. Let $\succeq$, U and W be as assumed.

(a) Given the assumptions, the argument is (informally) that if $p^W = q^W$, then p and q have the same 'welfare distribution', hence the same 'utility distribution' (as utility and welfare stand in one-to-one correspondence), and thus the same expected utility, which implies that $p \sim q$. Qed

(b) The order $\succeq^W$ is well-defined as the definition does not depend on the choice of p and q by (a). Qed

(c) Let ϕ be as specified. For all $p \in \mathcal{P}$, we have $\mathbb{E}_p U = \mathbb{E}_{p^W} \phi$, since

$$\begin{aligned}\mathbb{E}_p U &= \sum_{x \in X} p(x)U(x) = \sum_{w \in \mathbb{R}} \sum_{x \in X : W(x) = w} p(x)U(x) \\ &= \sum_{w \in \mathbb{R}} \left(\sum_{x \in X : W(x) = w} p(x) \right) \phi(w) \\ &= \sum_{w \in \mathbb{R}} p^W(w)\phi(w) = \mathbb{E}_{p^W} \phi.\end{aligned}$$

The claim now follows from the observation that, for any p^W and q^W in $\mathcal{P}^W$ (where $p, q \in \mathcal{P}$), $p^W \succeq^W q^W$ is equivalent to $p \succeq q$, hence to $\mathbb{E}_p U \geq \mathbb{E}_q U$, which reduces to $\mathbb{E}_{p^W} \phi \geq \mathbb{E}_{q^W} \phi$. Qed.

(d) First assume $\succeq^W$ displays constant classical risk aversion. To show CIRA, consider any $\Delta > 0$, any $p, p' \in \mathcal{P}$, and any $\rho, \rho' \in X$, such that $p \sim \rho$, $p' \sim \rho'$, and $p(W = w) = p'(W = w + \Delta)$ for each $w \in \mathbb{R}$. Then $p^W \sim^W \rho^W$, $p'^W \sim^W \rho'^W$, and $p^W(w) = p'^W(w + \Delta)$. So, as $\succeq^W$ displays constant classical risk aversion, $\rho'^W = \rho^W + \Delta$, i.e., $W(\rho') = W(\rho) + \Delta$. This establishes CIRA.

Conversely, assume CIRA. Consider any $\Delta > 0$, $a, a' \in \mathcal{P}^W$, and $t, t' \in Rg(W)$ such that $a \sim^W t$, $a' \sim^W t'$, and $a(w) = a'(w + \Delta)$ for each $w \in \mathbb{R}$ (where $a(w)$ stands for 0 if $w \notin Rg(W)$ and $a'(w + \Delta)$ stands for 0 if $w + \Delta \notin Rg(W)$). Pick $p, p' \in \mathcal{P}$ and $\rho, \rho' \in X$ such that $p^W = a$, $p'^W = a'$, $W(\rho) = t$ and $W(\rho') = t'$. Then $p \sim \rho$, $p' \sim \rho'$, and $p(W = w) = p'(W = w + \Delta)$ for each $w \in Rg(W)$. So, by CIRA, $W(\rho') = W(\rho) + \Delta$, i.e., $t' = t + \Delta$. This shows that $\succeq^W$ displays constant classical risk aversion. Qed

(e) Assume CIRA. The property established in (d) can be shown to imply that the risk premium has the same sign for all non-certain prospects, i.e., is always zero or always positive or always negative. This easily implies that U is linear or strictly concave or strictly convex, respectively. ■

The next lemma is a well-known building block of the classical theory of risk aversion after Arrow (1965) and Pratt (1964), and will later be applied to the order $\succeq^W$ in Lemma 2.

Lemma 3 *If an order on the set of finite-support lotteries over a given real interval has a smooth VNM representation with everywhere positive derivative, then it displays constant classical risk aversion if and only if it has a VNM representation given by* $w \mapsto \frac{1}{\rho}(e^{\rho w} - 1)$ *for some* $\rho \in \mathbb{R}$.

If $\rho = 0$, then '$\frac{1}{\rho}(e^{\rho w} - 1)$' of course stands for w $(= \lim_{\rho \to 0} \frac{1}{\rho}(e^{\rho w} - 1))$. Although this lemma is well-known, we sketch the argument for completeness.

Proof. Consider an order $\succeq^*$ on the set $\mathcal{P}^*$ of finite-support lotteries over a given interval $I \subseteq \mathbb{R}$, with a smooth VNM representation ϕ. For each $\rho \in \mathbb{R}$ let $\phi_\rho : I \to \mathbb{R}$ be the function $w \mapsto \frac{1}{\rho}(e^{\rho w} - 1)$. The proof goes in two steps.

Claim 1: $\succeq^*$ displays constant classical risk aversion if and only if there exists a $\rho \in \mathbb{R}$ such that ϕ solves the differential equation '$f'' = \rho f'$' on I, the solutions of which are the affine transformations of ϕ_ρ.

By the fundamental result of Arrow (1965) and Pratt (1964), $\succeq^*$ displays constant classical risk aversion if and only if the function ϕ''/ϕ' is constant, which implies the claimed 'if and only if'. The set of solutions to the differential equation '$f'' = \rho f'$' (on I) is well-known:

- If $\rho \neq 0$, then a solution is any affine transformation of the function $w \mapsto e^{\rho w}$.
- If $\rho = 0$, so that "$f'' = \rho f'$' reduces to '$f'' = 0$', then a solution is any affine transformation of the function $w \mapsto w$.

So, whatever the value of ρ, a solution of '$f'' = \rho f'$' is any affine transformation of ϕ_ρ. Qed.

Claim 2: If ϕ' is everywhere positive, then $\succeq^*$ displays constant classical risk aversion if and only if there exists a $\rho \in \mathbb{R}$ such that ϕ_ρ VNM represents $\succeq^*$.

Assume ϕ' is everywhere positive. Then ϕ and ϕ_ρ are two increasing functions, hence are increasing transformations of one another. By Claim 1, $\succeq^*$ displays constant classical risk aversion if and only if there exists a $\rho \in \mathbb{R}$ such that ϕ is a (now increasing) affine transformation of ϕ_ρ, or equivalently such that ϕ_ρ is an (increasing) affine transformation of ϕ, or yet equivalently such that ϕ_ρ (like ϕ) VNM represents $\succeq^*$. ■

Proof of Proposition 1. Consider any well-behaved $\succeq$ and W.

1. In this part we assume that $W = \log(\rho U + 1)/\rho$ for a VNM representation U of $\succeq$ and a $\rho \in \mathbb{R}$ such that $\rho U + 1 > 0$, and we prove that W satisfies CIRA. Note first that $U = (e^{\rho W} - 1)/\rho$. Thus, $U = \phi_\rho \circ W$, where ϕ_ρ is the function on $Rg(W)$ given

by $w \mapsto (e^{\rho w} - 1)/\rho$. Let $\succeq^W$ be the order over welfare prospects defined in Lemma 2. ϕ_ρ is a VNM representation of $\succeq^W$ by Lemma 2(c). So $\succeq^W$ displays constant classical risk aversion by Lemma 3. This implies CIRA by Lemma 2(d). Qed

2. Conversely, assume CIRA. We show the existence of a VNM representation U of $\succeq$ and a $\rho \in \mathbb{R}$ such that $\rho U + 1 > 0$ and $W = \log(\rho U + 1)/\rho$. Define $\succeq^W$ and the transformations $\phi_\rho : Rg(W) \to \mathbb{R}$ ($\rho \in \mathbb{R}$) as in part 1. Being well-behaved, $\succeq$ has a regular VNM representation $\tilde{U} : X \to \mathbb{R}$. As $\tilde{U}$ and W are regular and ordinally equivalent, $\tilde{U} = \phi \circ W$ for a smooth transformation $\phi : Rg(W) \to \mathbb{R}$ with $\phi' > 0$. ϕ VNM represents $\succeq^W$ by Lemma 2(c). CIRA implies that $\succeq^W$ displays constant classical risk aversion, by Lemma 2(d). Hence, by Lemma 3, there exists a $\rho \in \mathbb{R}$ such that ϕ_ρ VNM represents $\succeq^W$. As ϕ_ρ and ϕ both VNM represent $\succeq^W$, ϕ_ρ is an increasing affine transformation of ϕ. So, the function $U := \phi_\rho \circ W$ is an increasing affine transformation of $\tilde{U}$ $(= \phi \circ W)$. Hence, not only $\tilde{U}$ but also U VNM represents $\succeq$. Note that $\rho U + 1 > 0$, as $\rho U + 1 = \rho(\phi_\rho \circ W) + 1 > \rho(-1/\rho) + 1 = 0$. Finally, $W = \phi_\rho^{-1} \circ U = \log(\rho U + 1)/\rho$. ■

Proof of results in Section 3

Proof of Remark 1. As in Proposition 1, let the (well-behaved) functions U and W on X be related by $W = \log(\rho U + 1)/\rho$. So, $U = (e^{\rho W} - 1)/\rho$, and thus $U = \phi(W)$ where $\phi : Rg(W) \to \mathbb{R}$ maps any $w \in Rg(W)$ to $\phi(w) = (e^{\rho w} - 1)/\rho$ (which, as usual, reduces to w if $\rho = 0$). By simple computation, the intrinsic risk proneness ρ_W is given by

$$\rho_W = \frac{d^2U/dW^2}{dU/dW} = \frac{\phi''(W)}{\phi'(W)} = \frac{\rho^2 e^{\rho W}/\rho}{\rho e^{\rho W}/\rho} = \rho. \blacksquare$$

Proof of Proposition 2. Assume $X \subseteq \mathbb{R}$. Let $\succeq$ and W be well-behaved. As $\succeq$ is well-behaved, it has a regular VNM representation U. As W and U are ordinally equivalent and regular, the (unique) function $\phi : Rg(W) \to \mathbb{R}$ such that $U = \phi(W)$ is smooth with $\phi' > 0$. Differentiation yields

$$U' = \phi'(W)W' \text{ and } U'' = \phi''(W)W'^2 + \phi'(W)W''.$$

Hence the classical risk proneness $\rho_{AP} = \frac{U''}{U'}$ is given by

$$\rho_{AP} = \frac{\phi'(W)W'' + \phi''(W)W'^2}{\phi'(W)W'} = \frac{W''}{W'} + \frac{\phi''(W)}{\phi'(W)}W' = \frac{W''}{W'} + \rho_W W'. \blacksquare$$

Proof of Theorem 1

The proof of Theorem 1 will use Proposition as well as two further lemmas.

Lemma 4 *If $\succeq$ has a VNM representation U, then $\succeq$ is broad-ranging if and only if U is unbounded, i.e.,* $\sup U = \infty$ *or* $\inf U = -\infty$.

Proof. Assume $\succeq$ has a VNM representation U.

1. First let U be unbounded. Without loss of generality, suppose $\sup U = \infty$ (an analogous proof works if instead $\inf U = -\infty$). To prove that $\succeq$ is broad-ranging, consider situations $x, y \in X$ with $x \succeq y$. So $U(x) \geq U(y)$. As $\sup U = \infty$, there

is a situation $z \in X$ such that $U(z) - U(x) > U(x) - U(y)$. It easily follows that $\frac{1}{2}U(z) + \frac{1}{2}U(y) > U(x)$. So, as U VNM represents $\succeq$, $z_{\frac{1}{2}}y_{\frac{1}{2}} \succ x$.

2. Conversely, let $\succeq$ be broad-ranging. Then in particular not all situations in X are equally good. Pick any $x \succ y$ in X, and write $\Delta = U(x) - U(y)$ (> 0). For $j = 0, 1, ...$ define situations x_j and y_j with $U(x_j) - U(y_j) \geq 2^j\Delta$ by the following reclusion. First, $\bar{x} = x$ and $y_0 = y$. Clearly $U(\bar{x}) - U(y_0) \geq 2^0\Delta$ (in fact, '$\geq$' could be replaced by '$=$'). Now consider $j \geq 0$ and suppose x_j and y_j are defined, with $U(x_j) - U(y_j) \geq 2^j\Delta$. As $\succeq$ is broad-ranging, there exists a $g \in X$ such that $g_{\frac{1}{2}}y_{\frac{1}{2}} \succ x$ ('case 1') or there exists a $b \in X$ such that $y \succ b_{\frac{1}{2}}x_{\frac{1}{2}}$ ('case 2').

First assume case 1. Put $x_{j+1} = g$ and $y_{j+1} = y_j$. So, $\frac{1}{2}U(x_{j+1}) + \frac{1}{2}U(y_{j+1}) > U(x_j)$, and thus

$$\frac{1}{2}U(x_{j+1}) - \frac{1}{2}U(y_{j+1}) > U(x_j) - U(y_{j+1}) = U(x_j) - U(y_j) \geq 2^j\Delta.$$

Hence $U(x_{j+1}) - U(y_{j+1}) \geq 2^{j+1}\Delta$, as desired.

Now assume case 2 without case 1. Put $x_{j+1} = x_j$ and $y_{j+1} = b$. So, $\frac{1}{2}U(x_{j+1}) + \frac{1}{2}U(y_{j+1}) < U(y_j)$, and thus

$$\frac{1}{2}U(y_{j+1}) - \frac{1}{2}U(x_{j+1}) < U(y_j) - U(x_{j+1}) = U(y_j) - U(x_j) \leq 2^j\Delta.$$

Hence again $U(x_{j+1}) - U(y_{j+1}) \geq 2^{j+1}\Delta$, as desired.

As $j \to \infty$, we have $2^j\Delta \to \infty$, and so $U(x_j) - U(y_j) \to \infty$. So, $\sup U = \infty$ or $\inf U = -\infty$. ■

Lemma 5 *For any well-behaved observed order $\succeq$ and any welfare measure of the form $W = \log(\rho U + 1)/\rho$ for a VNM representation U of $\succeq$ and a $\rho \in \mathbb{R}$ such that $\rho U + 1 > 0$,*

(a) W satisfies Full-range if and only if $\inf U < 0 < \sup U$ and

$$\rho = \begin{cases} \frac{-1}{\sup U} \ (< 0) & \text{if } \sup U \neq \infty \\ \frac{-1}{\inf U} \ (> 0) & \text{if } \inf U \neq -\infty \\ 0 & \text{if } \sup U = \infty \text{ and } \inf U = -\infty, \end{cases}$$

assuming $\succeq$ is broad-ranging (so that U is unbounded by Lemma 4),

(b) W satisfies Normalisation if and only if U is normalised.

Proof. Let $\succeq$, U and ρ be as specified.

(a) Assume $\succeq$ is broad-ranging. So U is unbounded (Lemma 4). Since U is regular, $Rg(U)$ is an open interval. Thus $Rg(U) = (a, b)$ where we put $a = \inf U$ and $b = \sup U$. Note that $-\infty \leq a < b \leq \infty$, where at most one of a and b is finite.

If it is *not* the case that $a < 0 < b$, then $0 \notin Rg(U)$, and thus $0 \notin Rg(W)$; hence both sides of the claimed equivalence are violated, and thus the equivalence holds. Having set aside this trivial case, let us assume henceforth that $a < 0 < b$. As $Rg(U) = (a, b)$ and $W = \log(\rho U + 1)/\rho$, $Rg(W)$ is the open interval with the boundaries

$$\inf W = \lim_{u \downarrow a} \log(\rho u + 1)/\rho \text{ and } \sup W = \lim_{u \uparrow b} \log(\rho u + 1)/\rho.$$

This uses that $\log(\rho u+1)/\rho$ is a smooth and strictly increasing function of u, whether ρ is negative, positive, or zero (if $\rho = 0$ then $\log(\rho u + 1)/\rho$ stands for u, as usual). Since Full-range means that $Rg(W) = \mathbb{R}$, we have

$$\text{Full-range holds} \Leftrightarrow \quad \lim_{u\downarrow a} \log(\rho u + 1)/\rho = -\infty \text{ and } \lim_{u\uparrow b} \log(\rho u + 1)/\rho = \infty.$$

Thus, if ρ is positive, the claimed equivalence between Full-range and $\rho = -\frac{1}{a}$ holds because

$$\begin{aligned} \text{Full-range holds} &\Leftrightarrow \lim_{u\downarrow a} \log(\rho u + 1) = -\infty \text{ and } \lim_{u\uparrow b} \log(\rho u + 1) = \infty \\ &\Leftrightarrow \rho a + 1 = 0 \text{ and } \rho b + 1 = \infty \\ &\Leftrightarrow \rho = -\tfrac{1}{a} \text{ and } b = \infty \\ &\Leftrightarrow \rho = -\tfrac{1}{a}, \end{aligned}$$

where we could drop 'and $b = \infty$' since this is implied by a's finiteness and U's unboundedness. Analogously, if ρ is negative, then the claimed equivalence between Full-range and $\rho = -\frac{1}{b}$ holds because

$$\begin{aligned} \text{Full-range holds} &\Leftrightarrow \lim_{u\downarrow a} \log(\rho u + 1) = \infty \text{ and } \lim_{u\uparrow b} \log(\rho u + 1) = -\infty \\ &\Leftrightarrow \rho a + 1 = \infty \text{ and } \rho b + 1 = 0 \\ &\Leftrightarrow a = -\infty \text{ and } \rho = -\tfrac{1}{b} \\ &\Leftrightarrow \rho = -\tfrac{1}{b}. \end{aligned}$$

Finally, if $\rho = 0$, then the claimed equivalence between Full-range and $\rho = 0$ holds because the right side ($\rho = 0$) holds by assumption and the left side (Full-range) holds since $W = U$ and thus $Rg(W) = Rg(U) = \mathbb{R}$.

(b) We must show that W is normalised if only if U is normalised. This follows from two observations. First, as $W = \frac{1}{\rho}\log(\rho U + 1)$, W takes the value 0 exactly where U takes the value 0. Second,

$$W' = \frac{1}{\rho}\log'(\rho U + 1)(\rho U + 1)' = \frac{1}{\rho(\rho U + 1)}\rho U' = \frac{1}{\rho U + 1}U',$$

and so, wherever W (or equivalently U) takes the value 0, W' and U' coincide. ■

Proof of Theorem 1. Let $\succsim$ be well-behaved and broad-ranging. As $\succsim$ is well-behaved, it is VNM representable by a regular function. Note that any regular function has a normalised increasing affine transformation. So, there is a regular *and normalised* VNM representation U of $\succsim$. It is unbounded by Lemma 4. Note that $\inf U < 0 < \sup U$, because $Rg(U)$ is an open interval (by regularity) and contains 0 (by normalisation). So we may define

$$\rho = \begin{cases} \frac{-1}{\sup U} \ (< 0) & \text{if } \sup U \neq \infty \\ \frac{-1}{\inf U} \ (> 0) & \text{if } \inf U \neq -\infty \\ 0 & \text{if } \sup U = \infty \text{ and } \inf U = -\infty. \end{cases}$$

Further, we define the welfare measure $W = \log(\rho U+1)/\rho$. W is well-defined because $\rho U + 1 > 0$, by the definition of ρ and the fact that (since $Rg(U)$ is open) U is *strictly* larger than $\inf U$ and smaller than $\sup U$.

We now show that W is the only well-behaved welfare measure satisfying CIRA, Full-range and Normalisation. This will complete the proof, as it implies not only that there is a revealed measure, namely $W_{\succeq} = W$, but also that (by Remark 1) the revealed intrinsic risk proneness $\rho_{\succeq}$ is the constant ρ defined above.

Firstly, W is well-behaved as it is a smooth and positively differentiable transformation of the well-behaved function U. It also satisfies the three hypotheses: it satisfies CIRA by Proposition 1, and satisfies Full-range and Normalisation by Lemma 5.

Now let $\tilde{W}$ be any well-behaved welfare measure satisfying the hypotheses. We prove that $\tilde{W} = W$. By Proposition 1, CIRA implies that

$$\tilde{W} = \log(\tilde{\rho}\tilde{U} + 1)/\tilde{\rho}$$

for some VNM representation $\tilde{U}$ of $\succeq$ and some $\tilde{\rho} \in \mathbb{R}$ such that $\tilde{\rho}\tilde{U} + 1 > 0$. By Lemma 5 (applied with $\tilde{U}$ and $\tilde{\rho}$ rather than U and ρ), Normalisation implies that $\tilde{U} = U$, and Full-range implies that $\tilde{\rho} = \rho$ given that $\succeq$ is broad-ranging. So $\tilde{W} = W$. ■

Proof of Theorem 2 via Theorem 1

The following lemma will allow us to reduce Theorem 2 to Theorem 1.

Lemma 6 *For any observed order $\succeq$, any instance of the generalised conditions Full-range$_D$, Normalisation$_{r,s}$ and CIRA$_\tau$, any welfare measure $W : X \to \mathbb{R}$ such that $Rg(W) \subseteq D$ (so that $\tau \circ W$ is defined), and any increasing affine transformation W^* of $\tau \circ W$,*

- *(a) W is well-behaved if and only if W^* is well-behaved,*
- *(b) W satisfies Full-range$_D$ if and only if W^* satisfies Full-range,*
- *(c) W satisfies Normalisation$_{r,s}$ if and only if W^* satisfies Normalisation, assuming $W^* = (\tau \circ W - \tau(r))/(s\tau'(r))$,*
- *(d) W satisfies CIRA$_\tau$ if and only if W satisfies CIRA.*

Proof. Consider D, r, s, τ,W and W^* as specified. Let $\phi : D \to \mathbb{R}$ be the increasing affine transformation of τ such that $W^* = \phi \circ W$. Since τ is a smooth and positively differentiable function from D onto $\mathbb{R}$, so is ϕ. By basic analysis, it follows that ϕ^{-1} exists (so that $W = \phi^{-1} \circ W^*$) and that ϕ^{-1} is a smooth and positively differentiable function from $\mathbb{R}$ onto D.

(a) Recall that well-behavedness is the conjunction of compatibility with riskless comparisons and regularity. So the claim follows from two facts:

- W^* is compatible with riskless comparisons if and only if W is so, since W and W^* are ordinally equivalent.
- W^* is regular if and only if W is regular, since W and W^* are smooth positively differentiable transformations of one another.

(b) We have to show that $Rg(W) = D \Leftrightarrow Rg(W^*) = \mathbb{R}$. Note that $Rg(W^*) = \mathbb{R} \Leftrightarrow Rg(\tau \circ W) = \mathbb{R}$, as W^* is an increasing affine transformation of $\tau \circ W$. So it suffices to show that $Rg(W) = D \Leftrightarrow Rg(\tau \circ W) = \mathbb{R}$. This equivalence holds because, firstly, if $Rg(W) = D$ then $Rg(\tau \circ W) = \tau(Rg(W)) = \tau(D) = \mathbb{R}$, and secondly, if $Rg(W) \neq D$ then $Rg(\tau \circ W) = \tau(Rg(W)) \neq \tau(D) = \mathbb{R}$.

(c) Suppose $W^* = (\tau \circ W - \tau(r))/(s\tau'(r))$. In other words, $W^* = \phi \circ W$ where $\phi = (\tau(\cdot) - \tau(r))/(s\tau'(r))$. As a preliminary, consider the smooth transformation $\tilde{\phi}$ defined on $\tilde{D} = \{(d-r)/s : d \in D\}$ by $\tilde{\phi}(t) = \phi(st + r)$ for all $t \in \tilde{D}$. We have $\tilde{\phi}(0) = 0$ and $\tilde{\phi}'(0) = 1$, since $\tilde{\phi}(0) = \phi(r) = 0$ and $\tilde{\phi}'(t) = \frac{s\tau'(ts+r)}{s\tau'(r)}$ for all $t \in \tilde{D}$.

First, assume W satisfies Normalisation$_{r,s}$. Then $W(\overline{x}) = r$ and W has a derivative size of s at $\overline{x}$. Thus $W = s\tilde{W} + r$ for some normalised function $\tilde{W}$. Note that $\tilde{W} = (W - r)/s$ and that $Rg(\tilde{W}) = \{(t-r)/s : t \in D\} = \tilde{D}$. We have $\tilde{\phi} \circ \tilde{W} = W^*$, since

$$\tilde{\phi} \circ \tilde{W} = \tilde{\phi} \circ [(W - r)/s)] = \phi \circ W = W^*.$$

Since $\tilde{W}$ is normalised and since $W^* = \tilde{\phi} \circ \tilde{W}$ where $\tilde{\phi}$ is smooth with $\tilde{\phi}(0) = 0$ and $\tilde{\phi}'(0) = 1$, W^* is also normalised.

Conversely, assume W^* is normalised. Since ϕ is invertible, so is $\tilde{\phi}$ $(= \phi(s \times \cdot + r))$. Further, as $\tilde{\phi}$ is the composition of ϕ with the mapping $t \mapsto st + r$, ϕ^{-1} is the composition of the latter mapping with $\tilde{\phi}^{-1}$, i.e., $\phi^{-1} = s\tilde{\phi}^{-1}(\cdot) + r$. We thus have $W = \phi^{-1}(W^*) = s\tilde{\phi}^{-1}(W^*) + r$. To show that W satisfies Normalisation$_{r,s}$, it is thus sufficient to show that $\tilde{\phi}^{-1}(W^*)$ is normalised. This follows from the fact that W^* is normalised and the fact that $\tilde{\phi}^{-1}$ is smooth with $\tilde{\phi}^{-1} > 0$, $(\tilde{\phi}^{-1})(0) = 0$ and $(\tilde{\phi}^{-1})'(0) = 1$. The second fact holds because $\tilde{\phi}$ is smooth with $\tilde{\phi}' > 0$, $\tilde{\phi}(0) = 0$ and $\tilde{\phi}'(0) = 1$.

(d) By assumption, there are $a > 0$ and $b \in \mathbb{R}$ such that $W^* = a\tau(W) + b$, or equivalently $W = \tau^{-1}((W^* - b)/a)$.

First let W satisfy CIRA$_\tau$. To show that W^* satisfies CIRA, fix a $\Delta > 0$ and prospects $p, q \in \mathcal{P}$ with equivalent welfare w.r.t. W^* denoted w_p^* resp. w_q^*, and assume $p(W^* = w) = q(W^* = w + \Delta)$ for all $w \in \mathbb{R}$. We must show that $w_q^* = w_p^* + \Delta$. Since $W^* = a\tau(W) + b$, the prospects p and q have equivalent welfare w_p resp. w_q w.r.t. W satisfying $w_p^* = a\tau(w_p) + b$ resp. $w_q^* = a\tau(w_q) + b$. For all $t \in \mathbb{R}$, we have $p(\tau(W) = t) = q(\tau(W) = t + \Delta/a)$, because

$$p(\tau(W) = t) = p(a\tau(W) + b = at + b) = p(W^* = at + b)$$
$$q(\tau(W) = t + \Delta/a) = q(a\tau(W) + b = at + b + \Delta) = q(W^* = at + b + \Delta)$$

and because $p(W^* = w) = q(W^* = w + \Delta)$ for all $w \in \mathbb{R}$. We can now apply CIRA$_\tau$ to W and to Δ/a (rather than Δ). This yields $\tau(w_q) = \tau(w_p) + \Delta/a$. Thus $a\tau(w_q) + b = a\tau(w_p) + b + \Delta$, i.e., $w_q^* = w_p^* + \Delta$.

Conversely, suppose W^* satisfies CIRA. To show that W satisfies CIRA$_\tau$, note first that $Rg(W) \subseteq D$ by assumption. Next, consider a $\Delta > 0$ and prospects $p, q \in \mathcal{P}$ with equivalent welfare w_p resp. w_q, and assume $p(\tau(W) = t) = q(\tau(W) = t + \Delta)$ for all $t \in \mathbb{R}$. We prove that $\tau(w_q) = \tau(w_p) + \Delta$. As $W^* = a\tau(W) + b$, p and q have equivalent welfare w.r.t. W^* given by $w_p^* = a\tau(w_p) + b$ resp. $w_q^* = a\tau(w_q) + b$. For all $w \in \mathbb{R}$, we have $p(W^* = w) = q(W^* = w + a\Delta)$, because

$$p(W^* = w) = p(a\tau(W) + b = w) = p\left(\tau(W) = (w - b)/a\right)$$
$$q(W^* = w + a\Delta) = q(a\tau(W) + b = w + a\Delta) = q\left(\tau(W) = (w - b)/a + \Delta\right)$$

and because $p(\tau(W) = t) = q(\tau(W) = t + \Delta)$ for all $t \in \mathbb{R}$. So, by CIRA applied to W^* and to $a\Delta$ (rather than Δ), $w_q^* = w_p^* + a\Delta$, i.e., $a\tau(w_q) + b = a\tau(w_p) + b + a\Delta$. Thus, $\tau(w_q) = \tau(w_p) + \Delta$. ∎

Proof of Theorem 2. Assume $\succeq$ is well-behaved and broad-ranging, and consider the generalised hypotheses CIRA$_\tau$, Full-range$_D$ and Normalisation$_{r,s}$ for given D, τ, r, s.

1. In this part, we fix a well-behaved welfare measure $W : X \to \mathbb{R}$ satisfying the three hypotheses, and we prove that W has the specified form. By Lemma 6, the transformed welfare measure

$$W^* = (\tau \circ W - \tau(r))/(s\tau'(r)) \tag{1}$$

is well-behaved and satisfies the original hypotheses CIRA, Full-range and Normalisation. So, by Theorem 1, $W^* = \log(\rho_\succeq U + 1)/\rho_\succeq$, where U is the (unbounded) normalised VNM representation of $\succeq$, and $\rho_\succeq$ is the coefficient defined in Theorem 1. Defining ρ as in Theorem 2, and noting that $\rho_\succeq = \rho s\tau'(r)$, we obtain

$$W^* = \log(\rho s\tau'(r)U + 1)/(\rho s\tau'(r)). \tag{2}$$

By (1) and (2),

$$(\tau \circ W - \tau(r))/(s\tau'(r)) = \log(\rho s\tau'(r)U + 1)/(\rho s\tau'(r)).$$

Solving this equation for W yields $W = \tau^{-1}(\log(\rho s\tau'(r)U + 1)/\rho + \tau(r))$, as claimed. Qed.

2. In this part, we show that the welfare measure

$$W = \tau^{-1}(\log(\rho s\tau'(r)U + 1)/\rho + \tau(r)), \tag{3}$$

with U and ρ defined as in Theorem 2, is well-behaved and satisfies CIRA$_\tau$, Full-range$_D$ and Normalisation$_{r,s}$. By Lemma 6, this is the case if the transformed welfare measure W^* defined by (1) is well-behaved and satisfies CIRA, Full-range and Normalisation. By plugging the expression defining W into that defining W^*, and then simplifying, one obtains

$$W^* = \log(\rho s\tau'(r)U + 1)/(\rho s\tau'(r)) = \log(\rho_\succeq U + 1)/\rho_\succeq,$$

where $\rho_\succeq$ is defined like in Theorem 1, or equivalently $\rho_\succeq = \rho s\tau'(r)$. So, by Theorem 1, W^* is indeed well-behaved and satisfies CIRA, Full-range and Normalisation. ∎

References

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

Adler, M. D. (2016). Extended Preferences, in M.D. Adler, M. Fleurbaey (eds.), *Oxford Handbook of Well-Being and Public Policy*, Oxford University Press

Arrow, K. (1965) The theory of risk aversion. In: *Aspects of the Theory of Risk Bearing*. Helsinki: Yrjo Jahnssonin Saatio. Reprinted in: K. J. Arrow, *Essays in the Theory of Risk Bearing*. Amsterdam: North-Holland, 1971

Baccelli, J. (2018) Risk Attitudes in Axiomatic Decision Theory—A Conceptual Perspective, *Theory and Decision* 84: 61-82

Basu, K. (1982) Determinateness of the utility function: revisiting a controversy of the thirties, *Rev. Econom. Stud.* 49: 307–311

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

Bossert, W., Weymark, J. (2004) Utility in social choice. In: Barberà, S., Hammond P., Seidl, C. (eds.) *Handbook of utility theory, vol. 2: Extensions*, Kluwer, Dordrecht, pp. 1099-1177

Broome, J. (1991) *Weighing Goods*, Oxford: Blackwell

Buchak, L. (2013) *Risk and rationality*, Oxford University Press

Chiappori, P.-A., Paiella. M. (2011) Relative risk aversion is constant: evidence from panel data, *Journal of the European Economic Association* 9(6): 1021–1052

Dietrich, F., Jabarian, B. (2022) Decision under normative uncertainty, *Economics and Philosophy* 38(3): 372–394

Dhillon, A., Mertens, J.-F. (1999) Relative utilitarianism: An improved axiomatization, *Econometrica* 67: 471–498

Elminejada, A., Havraneka, T., Irsovaa, Z. (2022) Relative Risk Aversion: A Meta-Analysis, Working Paper, Charles University, Prague

Fleurbaey, M., Zuber, S. (2021) Fair Utilitarianism, American Economic Journal: Microeconomics 13(2): 370–401

Greaves, H. (2017) A reconsideration of the Harsanyi-Sen-Weymark debate on utilitarianism, *Utilitas* 29(2): 175–213

Pratt, J.W. (1964) Risk Aversion in the Small and in the Large, *Econometrica* 32: 122–136. https://doi.org/10.2307/1913738

Fleurbaey, M., Mongin, P. (2016) The utilitarian relevance of the aggregation theorem, *American Economic Journal: Microeconomics* 8(3): 289–306

Grant. S., Kajii, A., Polak, B., Safra, Z.. (2010) Generalized utilitarianism and Harsanyi's impartial observer theorem, *Econometrica* 78(6): 1939–71

Harsanyi, J. (1955) Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility, *J Political Economy* 63: 309–21

Harsanyi, J. (1978) Bayesian decision theory and utilitarian ethics, *American Economic Review* 68: 223-228

Isbell, J.R. (1959) Absolute games, in A.W. Tucker and R.D. Luce (eds.) *Contributions to the Theory of Games*, Vol. IV, Princeton University Press

Köbberling, V. (2006) Strength of preference and cardinal utility, *Econom. Theory* 27(2): 375–391

Krantz, D. H., Luce, R. D., Suppes, P., Tversky, A. (1971) Foundations of Measurement. In: *Additive and Polynomial Representations*, vol. 1, Academic Press, New York

McCarthy, D., Mikkola, K., Thomas, T. (2020) Utilitarianism with and without expected utility, *Journal of Mathematical Economics* 87: 77–113

Nissan-Rozen, I. (2015) Against Moral Hedging. *Economics and Philosophy* 31: 1-21

Pivato, M. (2013) Multiutility representations for incomplete difference preorders, *Mathematical Social Sciences* 66(3): 196–220

Segal, U. (2000) Let's agree that all dictatorship are equally bad, *Journal of Political Economy* 108: 569–589

Sen, A. (1977) Non-linear social welfare functions: A reply to Professor Harsanyi. In R. E. Butts & J. Hintikka (Eds.), Foundational problems in the special sciences, (Vol. 2, 297–302). Springer

Sen, A. (1985) *Commodites and Capabilities*. North-Holland, Amsterdam

Shapley, L. (1975) Cardinal utility comparisons from intensity comparisons, Tech. Rep. R-1683-PR, Rand Corporation

Stigler, G., Becker, G. (1977) De Gustibus non est disputandis, *American Economic Review* 67(2): 76–90

von Neumann, J., Morgenstern, O. (1944) *Theory of Games and Economic Behavior*, Princeton University Press

Wakker, P. (1988) The algebraic versus the topological approach to additive representations. *J. Math. Psych.* 32(4): 421–435

Wakker, P. (1989) *Additive Representations of Preferences*. Kluwer, Dordrecht.

Wakker, P. (2010) *Prospect Theory: for Risk and Ambiguity*, Cambridge University Press

Weymark, J. (1991) A reconsideration of the Harsanyi–Sen debate on utilitarianism. In Elster, J., Roemer, J. E. (eds.) *Interpersonal Comparisons of Well-Being*, Cambridge University Press, pp. 255