# Welfare vs. Utility

Franz Dietrich
Paris School of Economics & CNRS

Nov. 2022 / Jan. 2025[1]

**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. We prove that natural working hypotheses lead to a different welfare measure. It addresses familiar concerns about VNM utility, by faithfully capturing non-ordinal welfare features such as welfare intensity, despite resting on purely ordinal evidence such as revealed preferences or self-reported welfare comparisons. Using this welfare measure instead of VNM utility alters social welfare analysis – for instance, Harsanyi's 'utilitarian theorem' now effectively supports prioritarianism. VNM utility is shown to be a hybrid object, determined by an interplay of two factors: welfare and attitude to intrinsic risk, i.e., to risk in welfare rather than outcomes.

*Keywords*: welfare, utility, social welfare, utilitarianism, Harsanyi-Sen debate

## 1 Introduction

How should individual welfare be measured? This long-standing question matters throughout economics and applications. In practice, economists often measure welfare by VNM utility, largely because VNM utility is available based on ordinal evidence such as revealed preferences or self-reported comparisons between risky options. The notorious objection against this welfare measure is that it fails to adequately reflect non-ordinal welfare features such as welfare intensity, regardless of how one normalises VNM utility. Non-ordinal welfare features are however indispensable 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.

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

The controversy over whether VNM utility can measure welfare has culminated in the Harsanyi-Sen debate in the 1970s (later joined by Weymark 1991) and counts today among the most persistent points of divergence among welfare economists and formal ethicists. Critics of VNM utility as a welfare measure have a difficult standing so far, as they have not yet come up with an alternative measure based on ordinal evidence. The lack of ordinal foundations exposes these critics to the 'non-observability' objection – at least if one follows the ordinalist tradition that accepts only ordinal evidence. Ordinalism about evidence is itself controversial, but we will stick to this common assumption here (cf. Baccelli and Mongin 2016 on ordinalism and utility).

This paper provides a proof of concept for the VNM-sceptic position, by showing that purely ordinal 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 adopt the familiar idea, defended by Bell and Raiffa (1988) and Cibinel (2024), that someone's VNM utility function is influenced by two distinct features: her welfare or 'intrinsic utility' from outcomes, and her attitude to risk *in welfare* or 'intrinsic risk'. Unfortunately, one and the same VNM utility function can be given many different explanations in terms of welfare and intrinsic risk attitude – which *seems* to make welfare unobservable. Figure 1 indeed displays four rival explanations of the

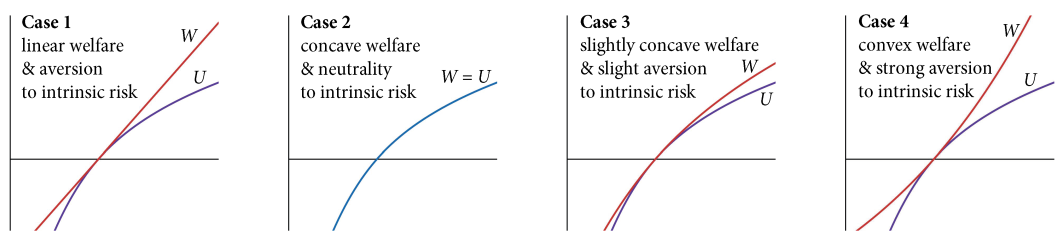


Figure 1: Four explanations of the same VNM utility function in terms of welfare and intrinsic risk attitude

same concave VNM utility function over wealth levels:[2]

- *Case 1: constant marginal welfare & aversion to intrinsic risk.* An extra 1\$ gives the same extra welfare regardless of initial wealth; so welfare is linear. Meanwhile a lottery with expected welfare $w$ is worse than a sure outcome with welfare $w$; so utility is concave *in welfare*, hence also concave in wealth, as welfare is linear. Note that we have invoked the intrinsic rather than standard risk attitude, by considering expected welfare, not expected wealth.

- *Case 2: diminishing marginal welfare & neutrality to intrinsic risk.* An extra \$1 gives less extra welfare if initial wealth is higher; so welfare is concave. Meanwhile a lottery with expected welfare $w$ is as good as a sure welfare of $w$;

[2] 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.

so utility coincides with welfare, hence is also concave.

- *Case 3: slightly diminishing marginal welfare & slight aversion to intrinsic risk.* An extra \$1 gives slightly less extra welfare if initial wealth is higher; so welfare is slightly concave. Meanwhile a lottery with expected welfare $w$ is slightly worse than a sure welfare of $w$; so utility is slightly concave in welfare, and thus concave in wealth.
- *Case 4: increasing marginal welfare & strong aversion to intrinsic risk.* An extra \$1 gives more extra welfare if initial wealth is higher; so welfare is convex. Meanwhile a lottery with expected welfare $w$ is much worse than a sure welfare of $w$; so utility is so strongly concave in welfare that it is also concave in wealth.

Seemingly, none of the four explanations can be ruled out empirically, which blocks the access to individual welfare, and thus blocks the possibility to assess social welfare, measure welfare inequality, or make welfare-based policy recommendations. It is therefore understandable that welfare economists are naturally drawn towards VNM utility as a proxy for welfare, at least when observability matters. However, the more the individuals depart from intrinsic risk neutrality, the less accurate this welfare proxy becomes for them – ultimately distorting social welfare judgments and welfare-based policies.

This paper introduces an new approach by which welfare *will* become observable, without drawing on non-ordinal evidence. But first, what does the literature say?

### Welfare and utility in the literature

Our view that VNM utility has two determinants – welfare (intrinsic utility) and attitude to risk in welfare (intrinsic risk attitude) – will resonate with many rational choice theorists, who routinely invoke one or the other or both determinants. Still, this picture can be challenged. We now sketch some prominent understandings of 'utility' and 'welfare', not all of which are compatible with our approach.

Sen (1977) and Weymark (1991) forcefully distinguish someone's welfare from her VNM utility. For them, someone's VNM utility functions have no privileged status: there are many other functions representing her order, some of which might better capture her welfare in a substantive, more-than-ordinal sense. We agree – and will add a concrete proposal of a welfare function, which will indeed not be of VNM-type.

Bell and Raiffa (1988), Nissan-Rosen (2015), Dietrich and Jabarian (2022) and many others endorse the distinction between welfare and VNM utility. Yet for instance Harsanyi, Broome (1991), 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. Yet his notion of welfare is not simply ordinal: for him, welfare intensity, levels and differences are meaningful concepts, captured by a suitably normalised VNM utility function. Note that, like us, he takes ordinal evidence (a VNM order) to generate a non-ordinal welfare measure

– but for him that measure is a VNM utility function. Unlike us, he does not distinguish between risk-attitudinal and welfare-based determinants of VNM utility: he treats risk aversion as a by-product of diminishing marginal welfare, not as a separate psychological disposition. Against this reductive view, Bell and Raiffa (1988) restore the conceptual independence between someone's risk-attitudinal and welfare-based features: it is perfectly possible to hate 'gambling' without having diminishing marginal welfare, or vice versa. We agree, and would add that Harsanyi's conflation of the two phenomena comes from cashing out risk attitudes as attitudes to risk in outcomes rather than welfare. This 'non-intrinsic' risk attitude is indeed influenced by diminishing marginal welfare – although it too should not be *reduced* to welfare features, being also influenced by something else, namely the intrinsic risk attitude, as will be shown formally.

First-generation economists used to think of utility in ways perfectly independent of any risk or risk attitude. Their term 'utility' corresponds to our 'welfare' or 'intrinsic utility', not to 'VNM utility'. Accordingly, their 'law of diminishing marginal utility' (from consumption) does not reflect risk aversion: it is effectively a 'law of diminishing marginal *welfare*'. Modulo terminology, their approach is in line with ours, but leaves open how to measure welfare from ordinal data – our central problem.

Arrow (1965) and Pratt (1964), the fathers of the modern theory of risk aversion, focus on VNM utility rather than welfare. They take someone's VNM utility function to be shaped entirely and solely by her risk attitude. At first, this seems to deny the idea that VNM utility has two determinants, welfare (intrinsic utility) and risk attitude. This impression however overlooks that Arrow-Pratt's 'risk attitude' is not the intrinsic risk attitude, but a hybrid object influenced by the intrinsic risk attitude and welfare, just like VNM utility. So, while their one-sided labels 'risk attitude' and 'theory of risk aversion' has caused some conceptual confusion by suppressing the role of welfare while stressing 'gambling taste', their theory stands in no formal conflict with our analysis. In passing, we will answer a question that they leave open: how can their measure of (non-intrinsic) risk aversion be reduced to its implicit determinants, intrinsic risk aversion and welfare? On classical risk attitudes, see also Baccelli (2018).

Peter Wakker (2010) takes an unorthodox, prospect-theoretic view, also endorsed by Abdellaoui et al. (2007) and Buchak (2013). This interesting view separates cleanly between welfare and risk attitude: while welfare affects VNM utility, the risk attitude only affects the weighting of probabilities, within a rank-dependent expected-utility model. Utility is thus explained by welfare alone – but not because the risk attitude is a mere by-product of marginal welfare (as for Harsanyi) but because this attitude leaves is mark elsewhere than in utility. By treating utility as a purely welfare-theoretic, not risk-attitudinal construct, the view conflicts with our two-determinant view on VNM utility, and trivialises our question of how to measure welfare – VNM utility itself is a measure. The view also conflicts with Arrow-Pratt's theory, which links VNM utility to the risk attitude – except under a radical reinterpretation of this theory, as a theory 'of diminishing marginal welfare' rather than 'of risk aversion'.

The measurement of welfare has also been studied extensively in *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), Nebel (forth.) 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 simply 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.[3] Like a VNM utility function, such a welfare measure is typically unique up to increasing affine transformation. But, unlike a VNM function, 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 ordinal evidence, where 'ordinal' refers to a binary order over options, not a difference order.[4]

## 2 Beyond Bernoulli: dropping intrinsic risk neutrality

Daniel Bernoulli (1738) famously postulated that one should maximise one's expected welfare, not one's expected monetary wealth. Like us, he used welfare as a primitive object – he called it 'utility', but it corresponds to our 'welfare' or 'intrinsic utility'. While he did not pursue our objective of making welfare observable, his welfare-based choice principle ties welfare closely to choice, thereby making welfare *partially observable*: welfare turns out to coincide with VNM utility, and is thus revealed by choice up to the two open parameters contained in VNM utility.

Bernoulli however makes a heavy implicit assumption: the agent is neutral to intrinsic risk, i.e., to risk in welfare. We will here drop this assumption, by allowing *any* intrinsic risk attitude, as long as this attitude is constant. As we show in this section, welfare then stays *partially observable* – with a third open parameter, which measures the intrinsic risk attitude. The question of how to determine the three parameters of welfare empirically will be postponed to Section 4.

We fix a set $X$ of *situations*, in which the welfare of a given individual is to be measured. While 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, wealth-health-education triples, or vectors of functionings (Sen 1985). 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 $k = 1$, so that $X$

[3] A function $W$ of the alternatives 'represents' a difference order $R$ over these alternatives if $(x, y)R(x', y') \Leftrightarrow W(x) - W(y) \geq W(x') - W(y')$ for all alternatives $x, y, x', y'$.

[4] Someone's choices still reveal an *incomplete fragment* of her difference order (Baccelli 2024).

is a non-empty open interval, e.g., $\mathbb{R}$, $(0, \infty)$ or $(0, 1)$.

A *welfare (or intrinsic utility) function* is a function $W : X \to \mathbb{R}$, where $W(x)$ represents the person's welfare in or from $x$. Welfare is not directly observed. Instead we observe ordinal comparisons between risky prospects, representing actions or policies. Let $\mathcal{P}$ be the set of *prospects*, i.e., lotteries over $X$ with finite support. The observable primitive is a binary relation $\succeq$ on $\mathcal{P}$, called a *prospect order*, where '$x \succeq y$' means that $x$ is observably at least as good as $y$ for the individual. The source of observation might be choice behaviour, self-assessments, third-party assessments, or perhaps neurophysiological data.

While we interpret $\succeq$ and $W$ mainly as capturing *welfare* comparisons resp. levels, one can replace the welfare-based interpretation by a preference-based interpretation, so that $\succeq$ and $W$ capture *preference* comparisons resp. preference strength. We will set aside the important welfare/preference distinction, and move back and forth freely between both interpretations – which is common, but deserves a special apology in the context of this paper.[5]

Let $\succ$ and $\sim$ denote the corresponding strict and asymmetric relations; '$x \succ y$' and '$x \sim y$' mean that $x$ is observably better than resp. as good as $y$ for the person.[6] Note that $X \subseteq \mathcal{P}$, identifying any situation $x \in X$ with the riskless prospect with sure outcome $x$.

How can we 'learn' $W$ from $\succeq$? The common move would identify welfare with VNM utility. A *VNM utility representation* of $\succeq$ is a function $U : X \to \mathbb{R}$ such that prospects are ranked by expected utility, i.e., for all prospects $p, q \in \mathcal{P}$, $p \succeq q \Leftrightarrow \mathbb{E}_p(U) \geq \mathbb{E}_q(U)$. If existent, such a representation is unique up to increasing affine transformation.

The common identification of welfare with VNM utility relies implicitly on the following hypothesis, which can be attributed to Daniel Bernoulli (1738), who said 'utility' for what we call 'welfare' or 'intrinsic utility':

**Intrinsic Risk Neutrality**: Any prospect is as good as a risk-free *welfare* level equal to the prospect's expected *welfare*. Formally, for all prospects $p \in \mathcal{P}$ and sure outcomes $x \in X$, if $\mathbb{E}_p(W) = W(x)$ then $p \sim x$.

By the following result, this hypothesis forces welfare to coincide with VNM utility, under a natural well-behavedness assumption. We call a function $W : X \to \mathbb{R}$ *regular* if it is smooth[7] with nowhere zero derivative $W'$, and *well-behaved* relative to $\succeq$ if it is moreover compatible with riskless comparisons, i.e., $W(x) \geq W(y) \Leftrightarrow x \succeq y$ for all riskless prospects $x, y \in X$.

[5] Note that while our label 'welfare function' for $W$ leans towards the welfare-based interpretation, our other label 'intrinsic utility function' stays neutral.

[6] 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 $q \succeq p$.

[7] 'Smooth' means that $W$ is differentiable arbitrarily many often. In the multi-dimensional case $X \subseteq \mathbb{R}^k$ with $k \geq 2$, $W'$ is of course a vector $(\frac{d}{dx_1}W, ..., \frac{d}{dx_k}W)$, and $W'$ is 'nowhere zero' if at no $x \in X$ it is the zero vector $(0, ..., 0)$.

**Proposition 1** *Given any prospect order $\succeq$, a well-behaved welfare function $W$ satisfies Intrinsic Risk Neutrality if and only if*

$$W = U$$

*for some VNM representation $U$ of $\succeq$.*

The Bernoullian hypothesis of Intrinsic Risk Neutrality was a significant progress at the time: it replaced the naive idea of neutrality to risk *in outcomes* with neutrality to risk *in welfare*.

Yet it is worrying (to say the least) that welfare measures given by VNM utility rest on a special attitude to intrinsic risk, namely neutrality. Bernoulli fixed only one of two problems: while he rightly made welfare the 'currency' of risk, he retained that special attitude to risk. We will replace the Bernoullian hypothesis by a more general hypothesis, which still locates risk at the welfare level, but now allows any attitude to risk – neutrality, aversion or proneness – as long as this attitude is stable, i.e., independent of the reference welfare level. Our generalised hypothesis relies on the *equivalent welfare* of a prospect $p \in \mathcal{P}$, defined as a welfare level $w_p$ achieved in a situation that is as good as $p$, formally $w_p = W(x_p)$ for some $x_p \in X$ such that $x_p \sim p$. This is our hypothesis (in its basic version, further generalised later):

**Constant Intrinsic Risk Attitude (CIRA)**: If a prospect is modified by a fixed increase *in welfare*, then the equivalent *welfare* increases by the same 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 $w_q = w_p + \Delta$.[8]

**Fact 1:** *CIRA generalises the Bernoullian hypothesis of Intrinsic Risk Neutrality.*

CIRA requires a 'coherent' or 'stable' attitude to intrinsic risk. For instance, if at low current welfare the agent is indifferent towards a 50:50 gamble that lets her gain 2 *welfare* units or lose 1 *welfare* unit, then she stays indifferent towards this gamble at higher welfare.

Just as Bernoulli's Intrinsic Risk Neutrality improves on an outcome-based concept of risk neutrality, CIRA improves on an outcome-based condition known as Constant Absolute Risk Aversion or 'CARA', as we will show soon. Much later, we will present a more systematic argument for CIRA, by deriving CIRA from two principles, namely dynamic consistency and welfare-change-based preferences (Section 5). Be this as it may, CIRA is not essential to our analysis, since our theorem can be generalised to a more flexible condition than CIRA (Section 5).

By being more general than the Bernoulli's Intrinsic Risk Neutrality, CIRA allows form a wider class of welfare functions. While Intrinsic Risk Neutrality forces one to

[8] Equivalently, *intrinsic* risk premia are invariant to welfare translations (assuming any 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 expected and equivalent welfare.

measure welfare by VNM utility (Proposition 1), CIRA implies the following welfare measurement. We will assume that the prospect order $\succeq$ is *regular*, i.e., has a regular VNM representation $U$.

**Proposition 2** *Given a regular prospect order $\succeq$, a well-behaved welfare function $W$ satisfies CIRA if and only if*

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

*for some VNM representation $U$ of $\succeq$ and some $\rho \in \mathbb{R}$ (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 case $\rho = 0$ is the special case of Intrinsic Risk Neutrality, treated in Proposition 1. In this sense, Proposition 2 generalises Proposition 1.

The parameter $\rho$ measures the attitude to intrinsic risk, i.e., to risk in welfare:

- if $\rho = 0$, the agent is *intrinsic risk neutral*, as VNM utility equals welfare,
- if $\rho > 0$, the agent is *intrinsic risk prone*, as VNM utility is convex in welfare (since welfare is concave in VNM utility)
- if $\rho < 0$, the agent is *intrinsic risk averse*, as VNM utility is concave in welfare (since welfare is convex in VNM utility).

Although the exact value of $\rho$ is usually empirically underdetermined, the sign of $\rho$ – and hence the *qualitative* intrinsic risk attitude – is often observable. This is so because $\rho$ must satisfy $\rho U + 1 > 0$. Specifically, the agent is observably

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

**Excursion: How CIRA can explain violations of CARA**

CIRA contrasts with a well-known condition, which operates at the level of outcomes rather than welfare, and rests on the standard notion of a prospect's certainty equivalent rather than equivalent welfare:[9]

**Constant Absolute Risk Aversion (CARA)**, defined for $X \subseteq \mathbb{R}$: If a prospect is modified by a fixed increase *in outcomes*, then the equivalent *outcome* increases by

[9] A certainty equivalent of $p \in \mathcal{P}$ is a situation $x_p \in X$ such that $p \sim x_p$.

the same 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 $x_q = x_p + \Delta$.[10]

CARA is implausible, as is confirmed by empirical violations (Chiappori and Paiella 2011). Why? If a risky wealth prospect is translated upwards, then it moves into a region of higher wealth and thus lower marginal welfare, assuming diminishing marginal welfare. So the new prospect contains less risk *in welfare*, i.e., less risk in a subjectively relevant sense; this leads to a smaller risk premium, assuming intrinsic risk aversion – in violation of CARA. For instance, 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: $W(10,000,000) \approx W(11,000,000)$. So the first lottery is equivalent to a wealth level close to the worse outcome \$0, the second to a wealth level close to the average outcome \$10,500,000, violating CARA.

CIRA can offer a systematic explanation for violations of CARA: CIRA rules out CARA, as long as welfare is not of a special form. To state this result, we call a real function on a subset of $\mathbb{R}$ *linear* if it is given by $x \mapsto ax + b$ for some $a, b \in \mathbb{R}$, *exponential* if it is given by $x \mapsto ae^{bx} + c$ for some $a, b, c \in \mathbb{R}$ with $a, b \neq 0$, and *logarithmic* if it is given by $x \mapsto a \log(bx + c)$ for some $a, b, c \in \mathbb{R}$ with $a, b \neq 0$.

**Fact 2** *(informal statement[11]): CIRA rules out CARA provided the welfare function is neither linear nor exponential nor logarithmic nor a logarithmic function of an exponential function.*

The natural Bernoullian response to empirical violations of CARA would be to make welfare the currency of risk, i.e., to replace CARA with CIRA. We follow this line. Modern choice theorists instead replace CARA with some other *outcome*-level hypothesis, such as 'hyperbolic absolute risk aversion' (HARA). This different response makes sense in light of the different objective: most choice theorists aim to represent or predict choice, while we aim to make welfare indirectly observable. The former objective precludes using conditions like CIRA that refer to unobservables, whereas our objective *requires* using conditions that relate the relevant unobservable ($W$) to the observable ($\succsim$), following the established scientific methodology for identifying unobservables.[12]

[10] To make '$p(x)$' and '$q(x + \Delta)$' well-defined even if $x$ resp. $x + \Delta$ fall outside $X$, 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 translations of outcomes (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 $\overline{p} = \sum_{x \in X} p(x)x$ and certainty equivalent $x_p$.

[11] The exact characterisation is stated in the appendix as Proposition 5.

[12] Applied economists, statisticians, psychologists, physicists and other empirical scientists all routinely rely on this approach: they all make inferences about relevant unobservables from observables via theoretic hypotheses linking the two. Of course, each field has its own type of observables, unobservables and hypotheses.

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

This section explores the structure of classic VNM utility and Arrow-Pratt risk attitude, by decomposing both quantities into their two fundamental determinants, welfare and intrinsic risk attitude. Our decompositions will show that standard utility and risk attitude are hybrid constructs, resulting from an interplay of two distinct ingredients. At this stage, the decomposition will still be partly unobservable, because welfare has still some open parameters in Proposition 2. Unique identification will be achieved later.

Arrow-Pratt's classic theory measures risk attitudes as follows, by assuming the one-dimensional case $X \subseteq \mathbb{R}$:

**Definition 1** *The* classical (or Arrow-Pratt) risk proneness *of a regular prospect order* $\succeq$*, for* $X \subseteq \mathbb{R}$*, is the (well-defined*[13]*) function* $\rho_{AP} = \rho_{AP,\succeq} = \frac{U''}{U'}$*, where* $U$ *is any VNM representation of* $\succeq$*. If constant,* $\rho_{AP}$ *is identified with its single value.*

One can measure the intrinsic risk attitude analogously, by simply replacing outcomes with welfare levels, which meanwhile allows us to lift the restriction to the one-dimensional case:

**Definition 2** *The* intrinsic risk proneness *of a regular prospect order* $\succeq$ *w.r.t. a well-behaved welfare function* $W$ *is the (well-defined*[14]*) function* $\rho_W = \rho_{W,\succeq} = \frac{d^2U/dW^2}{dU/dW}$*, where* $U$ *is any VNM representation of* $\succeq$*. If constant,* $\rho_W$ *is identified with its single value.*

The formula for $W$ in Proposition 2 implies that $U = (e^{\rho W} - 1)/\rho$, and hence that $\rho_W = \rho$, by simple algebra. So, Proposition 2 has two corollaries. First, we can replace $\rho$ with $\rho_W$ in Proposition 2:

**Corollary 1** *Given a regular prospect order* $\succeq$*, any well-behaved welfare function* $W$ *satisfying CIRA leads to a constant intrinsic risk proneness* $\rho_W$ *and takes the form* $W = \log(\rho_W U + 1)/\rho_W$*.*

Second, VNM utility can be explained by two determinants:

**Corollary 2** *Every regular prospect order* $\succeq$ *has a VNM utility representation* $U$ *determined by welfare and the intrinsic risk attitude via*

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

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

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

*for any well-behaved welfare function $W$ satisfying CIRA.*

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

Like VNM utility, the classical risk attitude can also be explained in terms of an interplay between welfare and the intrinsic risk attitude, this time without having to postulate CIRA:

**Proposition 3** *The classical risk attitude of any regular prospect order $\succeq$, with $X \subseteq \mathbb{R}$, is determined by welfare and the intrinsic risk attitude via*

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

*for any well-behaved welfare function $W$.*

Thus the classical risk proneness $\rho_{AP} = \frac{U''}{U'}$, i.e., 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 reflects the fact that risk in welfare matters only to the extent that welfare varies, i.e., to the extent $W'$.

As Proposition 3 does not require CIRA, the intrinsic risk attitude $\rho_W$ can be non-constant, and VNM utility $U$ need not equal $(e^{\rho_W W} - 1)/\rho_W$. Still $U$ must obey a differential equation: $\frac{U''}{U'} = \frac{W''}{W'} + W'\rho_W$. So VNM utility stays determined by welfare ($W$) and intrinsic risk attitude ($\rho_W$). Thus the central conceptual point – that classical utility has two distinct determinants – does not hinge on CIRA.

## 4 Uniquely revealed welfare and intrinsic risk attitude

So far, the welfare function $W$ is only partially revealed by the observable $\succeq$ – and this partial underdetermination is inherited by the intrinsic risk proneness $\rho_W$ and by the decompositions of VNM utility $U$ and classic risk proneness $\rho_{AP}$ into their two determinants $W$ and $\rho_W$. More precisely, the welfare function $W$ in Proposition 2 has three remaining degrees of freedom: $\rho$ 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 realistic 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 issues later, when we will generalise Full-range to allow for other measurement scales.

To normalise the welfare measure, we consider a fixed reference situation $\overline{x} \in X$, representing for instance a 'poverty point'. A function from $X$ to $\mathbb{R}$ (such as $W$) is *normalised* if at the reference point $\overline{x}$ 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.[15]

**Normalisation**: The welfare function $W$ is normalised.

Normalisation requires measuring welfare on a scale that sets welfare to 0 and marginal welfare size to 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 their intended meanings. One can always scale welfare such that Normalisation holds: every well-behaved welfare function satisfying CIRA and Full-range can be transformed into one that also satisfies Normalisation, by applying an increasing affine transformation. The fact that rescaling a welfare function changes welfare levels and welfare differences does not make welfare levels and differences meaningless. Rather it makes the meaning of levels and differences scale-relative: statements such as 'welfare is 2' and 'welfare rises by 3' can have meanings, which are fixed by the chosen scale. The axiom of Normalisation becomes less innocent when one engages in interpersonal comparisons of welfare levels and/or differences, because meanings cannot be fixed in more than one way. Later we will discuss this issue, and address it by generalising Normalisation.

We now state our central theorem, whereby there exists a unique, and therefore revealed, welfare function satisfying our hypotheses. It is obtained by choosing $U$ and $\rho$ in particular ways in the formula of Proposition 2. The result will assume that the prospect order $\succeq$ is *broad-ranging*. This means that for any situations there exist much better or much worse situations, more precisely: for any $x, y \in X$ with $x \succeq y$ there exists a $z \in X$ such that $z_{\frac{1}{2}}y_{\frac{1}{2}} \succ x$ or $y \succ z_{\frac{1}{2}}x_{\frac{1}{2}}$. Here, '$z_{\frac{1}{2}}y_{\frac{1}{2}} \succ x$ or $y \succ z_{\frac{1}{2}}x_{\frac{1}{2}}$' means that $z$ is *either* 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, including all HARA models.[16]

**Definition 3** *A prospect order* $\succeq$

[15] In the basic case $X \subseteq \mathbb{R}$, 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 $\|W'\|$.

[16] For all distinct $x, y \in X$ and all $t \in [0,1]$, $x_t y_{1-t}$ denotes the prospect $p \in \mathcal{P}$ such that $p(x) = t$ and $p(y) = 1 - t$.

- reveals welfare function $W$ *if $W$ is the <u>only</u> well-behaved welfare function satisfying CIRA, Full-range and Normalisation, in which case $W$ is denoted $W_{\succeq}$,*
- reveals intrinsic risk proneness $\rho$ *if $\succeq$ reveals welfare function $W_{\succeq}$ and $\rho = \rho_{W_{\succeq}}$, in which case $\rho$ is denoted $\rho_{\succeq}$.*

**Theorem 1** *Every regular and broad-ranging prospect order $\succeq$ reveals a welfare function and a constant intrinsic risk proneness, given by*

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

*and*

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

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

In practice, one should choose the VNM function $U$ to fit the data $\succeq$, and then derive $W_{\succeq}$ and $\rho_{\succeq}$ via Theorem 1. For instance, $U$ could be one of the many HARA utility functions, which often enjoy empirical confirmation, and are indeed unbounded.

The classical and intrinsic risk attitudes, $\rho_{AP}$ and $\rho_W$, can both be calculated from the VNM representation $U$, but in very different ways: while $\rho_{AP}$ $(= \frac{U''}{U'})$ is derived locally, from the curvature of $U$, $\rho_{\succeq}$ is derived globally, from the range of $U$.

# 5 Application, discussion and generalisation

In this section, we first sketch how our welfare measure can be combined with empirical data to yield concrete welfare assessments. We then turn to *social* welfare – the starting point of the Harsanyi-Sen debate about whether VNM utility measures welfare. This will then finally lead to a closer analysis of our three hypotheses, and to a generalisation of these hypotheses and the theorem.

## 5.1 Empirical application

Our formula for welfare (Theorem 1) can be applied using empirically supported VNM utility functions. For example, for $X = (0, \infty)$, assume the data support decreasing absolute risk aversion with a constant *relative* risk aversion of $\eta \geq 0$ ('CRRA'). Many studies confirm this picture, although the value of $\eta$ is highly context-dependent. The agent's normalised utility function is then of the well-known CRRA type:

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

If $\eta = 1$, this formula is interpreted as $U(x) = \overline{x} \log \frac{x}{\overline{x}}$ $(= \lim_{\eta \to 1} \frac{\overline{x}}{1-\eta} \left( \left(\frac{x}{\overline{x}}\right)^{1-\eta} - 1 \right))$. By Theorem 1, the agent's revealed intrinsic risk proneness is[17]

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

and her revealed welfare is[18]

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

So the CRRA model implies logarithmic welfare. Interestingly, welfare is independent of the relative risk aversion $\eta$. Thus the debate about the right value of $\eta$ does not affect welfare – it only affects VNM utility, hence choice. Here welfare is subject to less measurement uncertainty than VNM utility, despite being revealed more indirectly. Other models than the CRRA model – including other HARA models – lead to other concrete formulas for welfare and intrinsic risk proneness via Theorem 1.

### 5.2 Social welfare

Part of why measuring individual welfare matters is that it allows us to measure *social* welfare, as a guide to policy making. We will focus on *utilitarian* social welfare: social welfare is sum-total individual welfare. If John Harsanyi and followers are right, then individual welfare *is* VNM utility, and so 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 so one should maximise total welfare *rather than* total VNM utility. We now operationalise the latter view via Theorem 1.

Consider a society of individuals $i = 1, ..., n$ $(n \geq 2)$. Interpret $X$ as a set of social alternatives. Each individual $i$'s (well-behaved) prospect order $\succeq_i$ on $X$ reveals her welfare function $W_i = \log(\rho_i U_i + 1)/\rho_i$, where $\rho_i$ is her revealed intrinsic risk proneness and $U_i$ is her normalised VNM utility function. In many applications, $X = X_1 \times \cdots \times X_n$, where $X_i$ contains $i$'s possible situations (e.g., consumption bundles) and where $\succeq_i$ and $U_i$ are only sensitive to $X_i$, hence are essentially an order or function on $X_i$ rather than $X$.

What goes wrong when maximising total utility $\sum_i U_i$ (or total weighted utility[19]) rather than total welfare $\sum_i W_i$? Under the plausible assumption that individuals are intrinsic risk averse, each $U_i$ is a concave transformation of $W_i$. Thus, maximising total utility (or weighted utility) means maximising total *concavely transformed* welfare – an approach that prioritises the worse-off, and is known as *prioritarianism*, a famous alternative to utilitarianism (e.g., Adler 2019). So, ironically, the dedicated

[17] 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 $\inf U = -\infty$).

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

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

utilitarian John Harsanyi is effectively a prioritarian, and his 'utilitarian theorem' effectively supports prioritarianism.[20]

How should a utilitarian evaluate 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 maximise total ex-ante welfare $\sum_i \mathcal{W}_i(p)$, where $\mathcal{W}_i$ is the extension of $i$'s welfare function $W_i$ to $\mathcal{P}$ such that $\mathcal{W}_i(p) = W_i(x_p)$ for all $p \in \mathcal{P}$ with certainty equivalent $x_p \in X$. Either version of utilitarianism respects exactly one of the two conditions in Harsanyi's 'utilitarian' theorem: ex-post utilitarianism respects social VNM rationality, while ex-ante utilitarianism respects Pareto. Utilitarians thus face a hard choice in the face of risk, but both approaches can be operationalised using our welfare measure.

## 5.3 Generalisation

We now discuss possible objections to the three hypotheses, leading us to generalise the hypotheses and theorem. We begin with Full-range, followed by Normalisation, CIRA, and the theorem.

#### Full-range discussed and generalised

As measurement theorists will notice, Full-Range is not only a richness condition on $X$ (that forces $X$ to contain arbitrarily good or bad situations) but also implicitly a condition on the choice of measurement scale for welfare: that scale should have the range $\mathbb{R}$, so that all real numbers are *meaningful* as welfare levels. Scales with smaller range are also imaginable. For instance, a scale with range $\mathbb{R}$ can be transformed exponentially into one with range $\mathbb{R}_+$, by replacing ('relabelling') any welfare level $w \in \mathbb{R}$ with $e^w \in \mathbb{R}_+$. To allow many measurement scales, we fix a non-empty 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 $D = \mathbb{R}$):

**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 $\overline{x}$, for instance a poverty level. More generally, one can fix numbers $r \in D$ and $s > 0$, and place the following requirement, where we call a function from $X$ to $\mathbb{R}$ $(r, s)$-*normalised* if at the reference point $\overline{x}$ it takes the value $r$ and has a derivative of size $s$:

**Normalisation$_{r,s}$:** The welfare function $W$ is $(r, s)$-normalised.

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

What speaks for requiring Normalisation$_{r,s}$ instead of Normalisation, the special case with $r = 0$ and $s = 1$? Normalisation is questionable when one makes interpersonal comparisons of welfare, since Normalisation treats everyone as having the same welfare (and marginal welfare size) at $\overline{x}$. Nothing is wrong with assuming Normalisation for for a *given* person – this just requires choosing a measurement scale on which '0' and '1' have particular meanings adapted to that person (and such a scale always exists, as noted earlier). But assuming Normalisation for many persons simultaneously leads to questionable welfare comparisons at $\overline{x}$, since '0' and '1' can be given just one meaning at once. By contrast, Normalisation$_{r,s}$ works even in an interpersonal context, since $r$ and $s$ can be set differently for different persons.

In some contexts, Normalisation *is* defensible. Why? First of all, recall that choices of normalisation are an old problem in welfare economics, although it is usually raised for VNM utility rather than welfare. Already Harsanyi was bothered by the sensitivity of interpersonal utility comparisons and total utility in society to normalisation choices. Different proposals exist. Some fix utility at two reference outcomes, e.g., to 0 and 1 (e.g., Isbell 1959, Segal 2000, Adler 2012, 2016). Others fix the minimal and maximal utility (Karni and Weymark 2024). Fleurbaey and Zuber (2021) instead fix utility and marginal utility at a reference outcome. Our Normalisation follows their approach, applying it to welfare rather than VNM utility.

As we will explain, Normalisation is appropriate in three contexts:

*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, lying between absolute and ordinal information. Welfare levels and differences become meaningless, while welfare difference *ratios* stay significant – which creates a formal similarity between welfare and VNM utility, without erasing the fundamental difference.[21]

For affine welfare, Normalisation is unproblematic, since every welfare function is affinely equivalent to one satisfying Normalisation. While the utilitarian social welfare function $\sum_i W_i$ becomes meaningless, as it requires interpersonal comparisons of differences, other social welfare functions stay available, notably the Nash social welfare function, which requires only affine welfare.[22]

*2. Contextualised welfare*: Suppose the question is not what welfare the individuals *have* intrinsically, but what welfare they should be *treated* as having in a given social context, e.g., a context of welfare aggregation, commodity allocation, or policy choice.

[21] Welfare difference ratios and VNM utility difference ratios are both unique, but differ in value and meaning. While '$\frac{W(x)-W(y)}{W(x')-W(y')} = 2$' means that welfare changes twice as much from $x$ to $y$ as from $x'$ to $y'$, '$\frac{U(x)-U(y)}{U(x')-U(y')} = 4$' means something rather obscure, about a difference ratio for a hybrid object (VNM utility) combining welfare and intrinsic risk attitude.

[22] The latter is defined 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.

For such a contextualised notion of welfare, Normalisation says this: individuals should be *treated* as having identical welfare (of 0) and marginal welfare size (of 1) at the reference point $\overline{x}$. This sort of scaling has a remarkable implication: it ensures that Pigou-Dalton transfers increase the utilitarian social welfare. Let us be precise. Assume $X \subseteq \mathbb{R}$ and the reference point $\overline{x} \in X$ represents a 'poverty point' below/above which someone counts as poor/rich. Let social situations be vectors $(x_1, ..., x_n) \in X^n$, where $x_i$ represents $i$'s situation. Consider social welfare $\sum_i W_i(x_i)$ for $x \in X^n$. Then, once all $W_i$ satisfy Normalisation and are increasing and concave, social welfare $\sum_i W_i(x_i)$ increases by transferring resources from rich to poor persons.[23] So, Normalisation gives utilitarianism an unexpected egalitarian appeal. This egalitarian argument for Normalisation is introduced and developed axiomatically in Fleurbaey and Zuber (2021), in a version for VNM utility instead of welfare. In their words, Normalisation leads to 'fair utilitarianism'.

*3. Locally objective welfare*: Our analysis is open to 'objective' and 'subjective' notions of welfare (Fleurbaey and Blanchet 2013). Informally, objective welfare is determined by 'objective' features like wealth or consumption levels, subjective welfare by 'subjective' features like tastes for (or happiness from) wealth or consumption. As long as situations in $X$ are 'objective situations', i.e., represent the objective features, objective welfare is determined by the situation alone, while subjective welfare is also influenced by subjective features such as tastes. A notion of welfare can be hybrid, so that welfare depends partly on objective and partly on subjective features, where the extent of objectivity can vary across situations. One can interpret Normalisation as requiring welfare to be *objective (at least) at the reference point* $\overline{x}$: at $\overline{x}$, the person's welfare and marginal welfare size depend on objective features alone. For instance, a situation of misery $\overline{x}$ might lead *objectively* to a certain (low) well-being and (high) marginal welfare size, so that only non-miserable situations give room for subjective tastes to influence welfare. The slogan is: in situations of objective misery everyone is alike welfare-wise. Subjective features, such as whether one prefers Beethoven's or Bach's music, influence welfare only once extreme misery is overcome, i.e., once basic needs are satisfied. On this assumption, Normalisation is justified, subject to making the scaling convention that 0 (1) stands for the universal welfare (marginal welfare size) at $\overline{x}$.

In fact, it suffices that welfare be *effectively* locally objective at $\overline{x}$: at $\overline{x}$, welfare can still depend partly on subjective features as long as these features coincide for everyone at $\overline{x}$ – so that welfare at $\overline{x}$ depends on *subjective but universal* features. So, welfare can depend on subjective tastes, as long as everyone dislikes the situation of misery $\overline{x}$ *equally*, i.e., 'needs' subjectively the basic needs equally. This makes welfare effectively locally objective or 'locally intersubjective'.[24]

[23] That is, for all social situations $x, y \in X^n$ and individuals $j, k$, if $x_j < y_j < \overline{x} < y_k < x_k$, $y_j - x_j = x_k - y_k$, and $x_l = y_l$ for everyone else $l$, then $\sum_i W_i(x_i) < \sum_i W_i(y_i)$.

[24] Even if welfare is objective (really or effectively, locally or globally), people may differ in intrinsic risk attitudes, affecting their VNM utility functions.

The general idea is that the objective circumstances take over at $\overline{x}$, making subjective differences inexistent (true local objectivity) or irrelevant to well-being (effective local objectivity). The plausibility of this idea depends partly on the notion of situation in $X$. Mere wealth levels are perhaps too uninformative 'situations' for the welfare level to become objective at some 'poverty point' $\overline{x}$. Things might change if situations are detailed consumption vectors, or entire 'lives', or Sen-type functioning vectors. The more information is packed into situations – perhaps including quasi-subjective features – the less room is left for subjectivity in welfare assessments.

Stigler and Becker's (1977) thesis 'De gustibus non est disputandis' and Sen's (1985) programme of evaluating fine-grained functioning vectors are two very different attempts at objective evaluations through refining the description of situations.

### CIRA discussed and generalised

CIRA requires a constant attitude to intrinsic risk, i.e., to risk in welfare rather than outcomes – a welfare-level condition that can explain well-documented violations of the outcome-level condition CARA (see Fact 2 and Proposition 5).

One can defend CIRA by deriving it from two basic assumptions on how prospects are ordered. How? When facing a lottery $p \in \mathcal{P}$, the agent is in some status-quo situation, e.g., an initial wealth level. To make the status quo explicit, consider an entire *order structure*, i.e., a family $(\succeq_z)_{z \in X}$ containing the prospect order $\succeq_z$ held in any status quo $z \in X$. In a status quo $z \in X$, each lottery $p \in \mathcal{P}$ induces a *welfare-change lottery*, denoted $p_{\Delta W|z}$, defined as the finite-support lottery over $\mathbb{R}$ such that the probability of any welfare change $w \in \mathbb{R}$ is $p_{\Delta W|z}(w) = p(W(\cdot) - W(z) = w)$, the probability that final welfare minus initial welfare equals $w$. An order structure $(\succeq_z)_{z \in X}$ is

(a) *welfare-change-based* if, for all prospects $p, q, p', q' \in \mathcal{P}$ and stati quo $z, z' \in X$, if $p_{\Delta W|z} = p'_{\Delta W|z'}$ and $q_{\Delta W|z} = q'_{\Delta W|z'}$ then $p \succeq_z q \Leftrightarrow p' \succeq_{z'} q'$,

(b) *dynamically consistent* or *status-quo independent* if $\succeq_z$ is the same for each status quo $z \in X$.

**Proposition 4** *If an order structure $(\succeq_z)_{z \in X}$ satisfies (a) and (b), then the prospect order $\succeq_z \equiv \succeq$ satisfies CIRA, assuming it is regular and $W$ is well-behaved and satisfies Full-range.*

Condition (b) follows orthodox rational choice theory. Condition (a) follows Kahnemann-Tversky's popular approach of modelling decisions as choices between changes rather than final levels, but in an (arguably more plausible) version based on welfare changes rather than outcome changes. Kahneman and Tversky's idea is that real agents tend to conceptualise options in terms of changes rather than final consequences, since changes represent what 'happens' in the agent's perception.

Be this as it may, if one finds CIRA too restrictive, one can replace it with a more flexible hypothesis that can also accommodate other welfare-level hypotheses, such as *decreasing* intrinsic risk aversion, or constant *relative* intrinsic risk aversion. We define the generalised condition as requiring a constant attitude to risk in *some* welfare-based quantity, i.e., in some transformation of welfare. More precisely, given a welfare transformation, which can be any smooth function $\tau$ from the mentioned interval $D$ of meaningful welfare levels onto $\mathbb{R}$ with $\tau' > 0$, we require this:

**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 constant aversion to risk in welfare *ratios*, i.e., 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 regular and broad-ranging prospect order $\succeq$, there exists a unique well-behaved welfare function $W$ satisfying CIRA$_\tau$, Full-range$_D$ and Normalisation$_{r,s}$ (for given parameters $\tau, D, r, s$), namely the function*

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

*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 $\tau$ is the identity transformation, because the three hypotheses then reduce to the original ones, and the formulas reduce to those in Theorem 1.[25]

For instance, if $D = (0, \infty)$ and $\tau = \log$, so that CIRA$_\tau$ requires a constant *relative* intrinsic risk attitude, then the formula for $W$ reduces to $W = r(\frac{\rho s}{r} U + 1)^{1/\rho}$, so that welfare a geometric function of VNM utility.

## 6 Conclusion

We have shown that plausible working hypotheses allow one to operationalise the difficult notion of welfare or intrinsic utility. This operationalisation contrasts with

[25] If $\rho = 0$ then the expression $\log(\rho s \tau'(r) U + 1)/\rho$ in the formula for $W$ stands for $s\tau'(r)U$ ($= \lim_{\rho \to 0} \log(\rho s \tau'(r) U + 1)/\rho)$).

the popular one in terms of VNM utility – which rests on a less plausible working hypothesis, namely Intrinsic Risk Neutrality. Our approach allows one to decompose ordinary utility and risk attitude into its two determinants: welfare and intrinsic risk attitude. This makes welfare and intrinsic risk attitude indirectly observable, and suggests explaining the empirical finding of decreasing absolute risk aversion in terms of an interplay of decreasing marginal *welfare* and constant *intrinsic* risk attitude. We have given formal and informal reasons for adopting our hypotheses as working assumptions, but also presented generalised hypotheses that still make welfare observable, via a generalised formula.

Social welfare analysis can now use a more satisfactory observable measure of individual welfare than VNM utility, and it can disentangle welfare aspects from risk-attitudinal aspects, instead of both aspects being mixed unrecognisably within VNM utility. Critics of VNM utility in social welfare analysis, such as Amartya Sen and followers, so far failed to operationalise their position and to address the unobservability objection.

An rising challenge within social welfare theory is the social evaluation of risk – the ex-ante and ex-post approach represent two opposed views. A systematic analysis will ultimately need to spell out the *social* attitude to *intrinsic* risk. This could be achieved by aggregating not only individual into social welfare, but also individual into social intrinsic risk attitude. Since both individual characteristics – welfare and intrinsic risk attitude – are 'contained' in individual preferences under risk, this approach can be pursued within the standard framework, without introducing unobservables.

# A Appendix

## A.1 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 $k \geq 1$. Our definitions, hypotheses and results continue to hold for an arbitrary non-empty set $X$, except for those about the classical Arrow-Pratt risk attitude (Definition 1, Fact 2, and Propositions 3 and 5). This generalisation requires generalised notions of 'regular' and 'normalised' functions on $X$, since derivatives are undefined in general. Here are the details, for interested readers:

**Regularity generalised**. In the main text, a function $W : X \to \mathbb{R}$ counted as 'regular' if it is smooth with nowhere zero derivative. In general, the set of *regular* functions is 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 each strictly increasing $\phi : Rg(f) \to \mathbb{R}$, $\phi \circ f \in \mathcal{F} \Leftrightarrow \phi$ is smooth with $\phi' > 0$. The main text's regularity notion is a special case, as we will soon see.

**Normalisation generalised.** In the main text, a function $W : X \to \mathbb{R}$ counted as 'normalised' if it has value 0 and a derivative of size 1 at the reference point $\overline{x}$. In general, the set of *normalised* functions is any given set $\mathcal{N}$ of functions $f : X \to \mathbb{R}$ such that (i) for all $f \in \mathcal{N}$, $f(\overline{x}) = 0$, (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 function $f \in \mathcal{F}$ is normalisable, i.e., has an increasing affine transformation in $\mathcal{N}$. Normalised functions have the right value at $\overline{x}$ by (i), and intuitively the same 'abstract derivative' at $\overline{x}$ by (ii).[26] We must also generalise the related notion of an $(r, s)$-*normalised* function $f : X \to \mathbb{R}$, for $r \in \mathbb{R}$ and $s > 0$. In the main text, '$(r, s)$-normalised' means that $f(\overline{x}) = r$ and $f$ has a derivative of size $s$ at $\overline{x}$. In general, it means that $f = sg + r$ for some normalised function $g \in \mathcal{N}$.

The concrete setup of the main text is indeed a special case:

**Lemma 1** *The conditions on sets $\mathcal{F}$ and $\mathcal{N}$ of regular resp. normalised functions hold if (as in the main text) $X$ is a non-empty open connected subset of $\mathbb{R}^k$ with $k \geq 1$ and*

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

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

1. Fix an $f \in \mathcal{F}$. $Rg(f)$ is an interval, since continuous images of connected sets are connected. This interval is open, because $X$ is open and $f'(x) \neq \mathbf{0}$ for all $x \in X$. Now fix a strictly increasing $\phi : Rg(f) \to \mathbb{R}$. By basic calculus, if $\phi$ is smooth with $\phi' > 0$, then $\phi \circ f \in \mathcal{F}$.

Conversely, assume $\phi \circ f \in \mathcal{F}$. We must show that $\phi$ 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 $n^{\text{th}}$ 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 takes the claimed form, say $h^{(n-1)} = \frac{a}{b}$. By implication, $h^{(n)}$ exists and equals $\frac{a'b - b'a}{b^2}$, another ratio of the claimed form. Q.e.d..

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

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 $\phi$ can be deduced from the fact that $\phi = g \circ h$ and

[26] In (ii), $\phi \circ f$ intuitively has the same abstract derivative at $\overline{x}$ as $f$ if and only if $\phi'(0) = 1$. Intuitive 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.

that $g$ and (by Claim 1) $h$ are smooth. How? In short, $\phi'$ exists and equals $h'g'(h)$; so $\phi''$ exists and equals $h''g'(h) + h'(g'(h))' = h''g'(h) + h'^2 g''(h)$; and so on for higher derivatives of $\phi$ (we skip the full inductive argument).

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 $\phi'(w) > 0$. Q.e.d.

*Claim 3*: In general, $\phi$ is smooth with $\phi' > 0$ (completing the proof).

Now we allow $X$ to be multi-dimensional: $X$ is any non-empty open connected subset of $\mathbb{R}^k$ with $k \geq 1$. Let $t \in Rg(f)$. We must show that, at $t$, $\phi$ 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 $j^{\text{th}}$ coordinate in $\tilde{X}$. That is, consider the function $\tilde{f} : \tilde{X} \to \mathbb{R}$, $y \mapsto f(x_1, ..., x_{j-1}, y, x_{j+1}, ..., x_n)$. Let $\tilde{\phi}$ be the restriction of $\phi$ to $Rg(\tilde{f})$ ($\subseteq Rg(f)$). We now replace the primitives $X$, $f$, $\phi$ and $\mathcal{F}$ with, respectively, $\tilde{X}$, $\tilde{f}$, $\tilde{\phi}$ and $\tilde{\mathcal{F}} = \{s : \tilde{X} \to \mathbb{R} : s \text{ is smooth } \& \ s'(x) \neq 0$ for all $x \in \tilde{X}\}$. Note that we indeed have $\tilde{f} \in \tilde{\mathcal{F}}$ (show this using that $f \in \mathcal{F}$) and $\tilde{\phi} \circ \tilde{f} \in \tilde{\mathcal{F}}$ (show this 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 $\phi$ coincides with $\tilde{\phi}$ on $Rg(\tilde{f})$, $\phi$ is smooth with $\phi' > 0$ on $Rg(\tilde{f})$, hence at $t$.

2. We now 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 $\phi \circ f \in \mathcal{N}$, since $\phi \circ f(\overline{x}) = \phi(0) = 0$, and since $(\phi \circ f)'(\overline{x})$ exists (as $f'(\overline{x})$ and $\phi'$ exist) with $\|(\phi \circ f)'(\overline{x})\| = \|\phi'(f(\overline{x}))f'(\overline{x})\| = |\phi'(f(\overline{x}))| \, \|f'(\overline{x})\| = 1 \times 1 = 1$. 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 $\|g'(\overline{x})\| = \frac{1}{\|f'(\overline{x})\|} \|f'(\overline{x})\| = 1$. ■

## A.2 Proof of all propositions and Fact 2

Depending on one's taste, one can read the following proofs either with the main text's concrete setup in mind or with the generalised setup in mind – except of course for the few results about the classic risk attitude, which hold for the concrete setup only.

*Proof of Proposition 1.* Fix a prospect order $\succeq$ and a well-behaved welfare function $W$. First, if $W$ VNM represents $\succeq$, i.e., if $\succeq$ ranks prospects by expected welfare, then Intrinsic Risk Neutrality holds obviously. Conversely, assume Intrinsic Risk Neutrality. We must show that $W$ VNM represents $\succeq$. We fix $p, q \in \mathcal{P}$ and will prove that $p \succeq q \Leftrightarrow \mathbb{E}_p(W) \geq \mathbb{E}_q(W)$. As $W$ is regular, $Rg(W)$ is an interval. So, $\mathbb{E}_p(W), \mathbb{E}_q(W) \in Rg(W)$, and thus there exist $x, y \in X$ such that $W(x) = \mathbb{E}_p(W)$ and $W(y) = \mathbb{E}_q(W)$. By Intrinsic Risk Neutrality, $x \sim p$ and $y \sim q$. So, $p \succeq q$ is equivalent to $x \succeq y$, hence to $W(x) \geq W(x)$ (as $W$ is compatible with riskless comparisons), i.e., to $\mathbb{E}_p(W) \geq \mathbb{E}_q(W)$, as desired. ■

Some notation and lemmas will prepare our next proofs.

*Notation.* Given a welfare function $W$, 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 $p^W$, where for each $w \in W$ we define $p^W(w)$ as $p(W = w)$ $(= p(\{x \in X : W(x) = w\}))$.

**Lemma 2** *Assume $\succeq$ has a VNM representation $U$, and $W : X \to \mathbb{R}$ is ordinally equivalent to $U$. 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) $p, q \in \mathcal{P}$ such that $p^W = a$ and $q^W = b$.*
- *(c) $\succeq^W$ has a VNM representation given by the (strictly increasing) function $\phi : Rg(W) \to \mathbb{R}$ such that $U = \phi \circ W$.*
- *(d) $\succeq^W$ satisfies CARA if and only if $\succeq$ and $W$ satisfy CIRA.*
- *(e) In particular, if CIRA holds, $\phi$ 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 $p$ and $q$ have the same welfare prospect (i.e., $p^W = q^W$), then they have the same 'utility prospect' (as utility is a one-to-one function of welfare), and hence the same expected utility, implying that $p \sim q$. Q.e.d.

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

(c) Let $\phi$ 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 $\mathbb{E}_{p^W}(\phi) \geq \mathbb{E}_{q^W}(\phi)$. Q.e.d..

(d) First assume $\succeq^W$ satisfies CARA. 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 $\succeq^W$ satisfies CARA, $\rho'^W = \rho^W + \Delta$, 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 all $w \in \mathbb{R}$ (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 $\succeq^W$ satisfies CARA. Q.e.d.

(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 satisfies CARA 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 $\phi$. 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*: $\succeq^*$ satisfies CARA if and only if there exists a $\rho \in \mathbb{R}$ such that $\phi$ solves the differential equation '$f'' = \rho f'$' on $I$, the solutions of which are the affine transformations of $\phi_\rho$.

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

- If $\rho \neq 0$, then the solutions are the affine transformations of the function $w \mapsto e^{\rho w}$.
- If $\rho = 0$, so that "$f'' = \rho f'$' reduces to '$f'' = 0$', then the solutions are the affine transformations of the function $w \mapsto w$.

So, whether $\rho \neq$ or $\rho = 0$, the solutions are the affine transformations of $\phi_\rho$. Q.e.d..

*Claim 2*: If $\phi'$ is everywhere positive, then $\succeq^*$ satisfies CARA if and only if there exists a $\rho \in \mathbb{R}$ such that $\phi_\rho$ VNM represents $\succeq^*$.

Assume $\phi'$ is everywhere positive. Then $\phi$ and $\phi_\rho$ are two increasing functions, hence are increasing transformations of one another. By Claim 1, $\succeq^*$ satisfies CARA if and only if there exists a $\rho \in \mathbb{R}$ such that $\phi_\rho$ and $\phi$ are (now increasing) affine transformations of one another, or equivalently such that $\phi_\rho$ (like $\phi$) VNM represents $\succeq^*$. ■

*Proof of Proposition* 2. Consider any regular $\succeq$ and well-behaved $W$.

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 $\phi_\rho$ 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. By Lemma 2(c), $\phi_\rho$ VNM represents $\succeq^W$. So $\succeq^W$ satisfies CARA by Lemma 3. This implies CIRA by Lemma 2(d). Q.e.d.

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 before. Being regular, $\succeq$ 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 $\phi' > 0$. $\phi$ VNM represents $\succeq^W$ by Lemma 2(c). CIRA implies that $\succeq^W$ satisfies CARA by Lemma 2(d). Hence, by Lemma 3, there exists a $\rho \in \mathbb{R}$ such that $\phi_\rho$ VNM represents $\succeq^W$. As $\phi_\rho$ and $\phi$ both VNM represent $\succeq^W$, $\phi_\rho$ is an increasing affine transformation of $\phi$. 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 $\succeq$. We have $\rho U + 1 > 0$, 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 Proposition 3.* Assume $X \subseteq \mathbb{R}$. Let $\succeq$ be regular and $W$ well-behaved. As $\succeq$ is regular, 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

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

*Proof of Proposition 4.* Assume an order structure $(\succeq_z)$ satisfying (a) and (b), and let $\succeq \equiv \succeq_z$ be regular and $W$ well-behaved. Let $\Delta, p, w_p, q, w_q$ be as in CIRA. Let $p(W = w) = q(W = w + \Delta)$ for all $w \in \mathbb{R}$. We must show that $w_q = w_p + \Delta$. By (a), there exists an order $\succeq^*$ over welfare-change lotteries (i.e., finite-support lotteries on $\mathbb{R}$) such that $\tilde{p} \succeq_{\tilde{z}} \tilde{q} \Leftrightarrow \tilde{p}_{\Delta W|\tilde{z}} \succeq^* \tilde{q}_{\Delta W|\tilde{z}}$ for all $\tilde{p}, \tilde{q} \in \mathcal{P}$ and $\tilde{z} \in X$. Pick situations $z, z' \in X$ such that $W(z') = W(z) + \Delta$ (they exist by Full-range). Pick certainty equivalents $x_p, x_q \in X$ of $p$ resp. $q$ (they exist as $\succeq$ is regular). Now $p \sim x_p$, and so $p \sim_z x_p$, whence $p_{\Delta W|z} \sim^* W(x_p) - W(z) = w_p - W(x_0)$ (identifying the welfare change $W(x_p) - W(z) = w_p - W(x_0)$ with a riskless welfare-change lottery). Analogously, $q \sim x_q$, and so $q \sim_{z'} y$, whence $q_{\Delta W|z'} \sim^* W(x_q) - W(z') = w_q - W(z')$. Further, $p_{\Delta W|z} = q_{\Delta W|z'}$, since $q$'s welfare prospect equals $p$'s shifted by $\Delta$, and $z'$'s welfare equals $z$'s shifted by $\Delta$. In sum, $p_{\Delta W|z} \sim^* w_p - W(z)$, $q_{\Delta W|z'} \sim^* w_q - W(z')$, and $p_{\Delta W|z} = q_{\Delta W|z'}$. Thus $w_p - W(z) \sim^* w_q - W(z')$. So $w_p - W(z) = w_q - W(z')$, since indifferent welfare changes are identical (as $\succeq$ is regular and $W$ is well-behaved). Thus, $w_q - w_p = W(z') - W(z) = \Delta$. ■

We now formally restate and prove Fact 2, which claims that CIRA can explain violations of CARA whenever the welfare function is not of a special type. More precisely:

**Fact 2** *(formal statement): For any regular prospect order $\succeq$ with $X \subseteq \mathbb{R}$, CARA is violated if CIRA holds for a well-behaved welfare function $W$ that is neither linear nor exponential nor logarithmic nor a logarithmic function of an exponential function.*

Fact 2 follows from a more general characterisation:

**Proposition 5** *Given a regular prospect order $\succeq$ and a well-behaved welfare function $W$ satisfying CIRA, where $X \subseteq \mathbb{R}$, CARA holds if and only if $W$ is (i) a linear or exponential function with $\rho_W = 0$, or (ii) the base-$e^{\rho_W}$ logarithm of such a function with $\rho_W \neq 0$.*

For instance, if $\rho_W = -1$ (a form of intrinsic risk aversion), then CARA holds precisely if welfare takes the form $W = \frac{\log(V)}{\log(e^{-1})} = -\log(V)$ for some linear or exponential function $V$.

The proof of Proposition 5 will draw on the following well-known result, which is an obvious variant of Lemma 3 (without 'positive derivative' restriction):

**Lemma 4** *If an order on the set of finite-support lotteries over a given real interval has a smooth VNM representation $U$, then it satisfies CARA if and only if $U$ is linear or exponential.*

*Proof of Proposition 5.* Assume $X \subseteq \mathbb{R}$, $\succeq$ is regular, $W$ is well-behaved, and CIRA holds. By CIRA, $\rho_W$ is constant (see Corollary 1).

1. First, assume CARA. We show that $W$ takes one of the special forms. CIRA implies that there is a VNM representation $U$ of $\succeq$ such that $W = \log(\rho_W U + 1)/\rho_W$, interpreted as $W = U$ if $\rho_W = 0$ (Proposition 2 and Corollary 1). Meanwhile CARA

implies that $U$ is linear or exponential (Lemma 4). So, if $\rho_W = 0$, then $W$ $(= U)$ is linear or exponential, while if $\rho_W \neq 0$, then

$$W = \log(\rho_W U + 1)/\rho_W = \log(\rho_W U + 1)/\log(e^{\rho_W}),$$

i.e., $W$ is the base-$e^{\rho_W}$ logarithm of a function that is linear (if $U$ is linear) or exponential (if $U$ is exponential). Thus $W$ takes one of the special forms.

2. Now assume $W$ takes one of these forms. By CIRA, $\succeq$ is VNM representable by a function $U$ that we can scale such that $U = W$ if $\rho_W = 0$ and $U = ke^{\rho_W W}$ for a $k \in \mathbb{R}$ of same sign as $\rho_W$ if $\rho_W \neq 0$ (see Corollary 2). First assume $\rho_W = 0$. Then $U = W$, and $W$ is linear or exponential. So $U$ is linear or exponential, implying CARA (Lemma 4).

Now assume $\rho_W \neq 0$. Then $U = ke^{\rho_W W}$, and $W = \frac{1}{\rho_W}\log V$ for a linear or exponential function $V$. Thus

$$U = e^{\rho_W W} = e^{\rho_W \frac{1}{\rho_W}\log V} = e^{\log V} = V.$$

So $U$ is linear or exponential, again implying CARA (Lemma 4). ■

## A.3 Proof of Theorem 1

The proof of Theorem 1 will use Proposition 2 as well as two further lemmas.

**Lemma 5** *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. 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 $U(x_j) - U(y_j) \geq 2^j\Delta$ recursively as follows. First, $x_0 = x$ and $y_0 = y$. Clearly $U(\overline{x}) - U(y_0) \geq 2^0\Delta$ (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 but not 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 6** *For any regular prospect order $\succeq$ and any welfare function 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 5),*

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

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

(a) Assume $\succeq$ is broad-ranging. So $U$ is unbounded (Lemma 5). Since $U$ is regular, $Rg(U)$ is an open interval. Thus $Rg(U) = (a, b)$ where $a = \inf U$ 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. Now assume 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 $\rho$ 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 $\rho$ is positive, the claimed equivalence between Full-range and $\rho = -\frac{1}{a}$ holds because

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

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

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

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,

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

so that $W'$ and $U'$ coincide wherever $W$ (or equivalently $U$) is 0. ■

*Proof of Theorem 1.* Let $\succeq$ be regular and broad-ranging. As $\succeq$ is regular, 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 $\succeq$. It is unbounded by Lemma 5. 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 function $W = \log(\rho U + 1)/\rho$. $W$ is well-defined because $\rho U + 1 > 0$, by the definition of $\rho$ 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 function 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 Corollary 1) the revealed intrinsic risk proneness $\rho_{\succeq}$ is the constant $\rho$ defined above.

Firstly, $W$ is well-behaved and satisfies the hypotheses: well-bahavedness holds since $W$ is a smooth and positively differentiable transformation of the well-behaved function $U$, CIRA holds by Proposition 2, and Full-range and Normalisation hold by Lemma 6.

Secondly, let $\tilde{W}$ be *any* well-behaved welfare function satisfying the hypotheses. We prove that $\tilde{W} = W$. By Proposition 2, CIRA implies $\tilde{W} = \log(\tilde{\rho}\tilde{U} + 1)/\tilde{\rho}$ 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 6 (applied with $\tilde{U}$ and $\tilde{\rho}$ rather than $U$ and $\rho$), Normalisation implies $\tilde{U} = U$, and Full-range implies $\tilde{\rho} = \rho$ given that $\succeq$ is broad-ranging. So $\tilde{W} = W$. ■

## A.4 Proof of Theorem 2 via Theorem 1

The following lemma prepares the reduction of Theorem 2 to Theorem 1.

**Lemma 7** *For any prospect order $\succeq$, any instance of the generalised conditions Full-range$_D$, Normalisation$_{r,s}$ and CIRA$_\tau$, any welfare function $W$ such that $Rg(W) \subseteq D$ (ensuring 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 $D$, $r$, $s$, $\tau$,$W$ and $W^*$ as specified. Let $\phi : D \to \mathbb{R}$ be the increasing affine transformation of $\tau$ such that $W^* = \phi \circ W$. Since $\tau$ is a smooth and positively differentiable function from $D$ onto $\mathbb{R}$, so is $\phi$. By basic analysis, it follows that $\phi^{-1}$ exists (so that $W = \phi^{-1} \circ W^*$) and that $\phi^{-1}$ 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 $W = s\tilde{W} + r$ for some normalised $\tilde{W}$. 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 $\phi$ is invertible, so is $\tilde{\phi}$ $(= \phi(s \times \cdot + r))$. Further, as $\tilde{\phi}$ is the composition of $\phi$ with the mapping $t \mapsto st + r$, $\phi^{-1}$ 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

$$\begin{aligned}
&p(\tau(W) = t) = p(a\tau(W) + b = at + b) = p(W^* = at + b) \\
&q(\tau(W) = t + \Delta/a) = q(a\tau(W) + b = at + b + \Delta) = q(W^* = at + b + \Delta)
\end{aligned}$$

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 $\Delta/a$ (rather than $\Delta$). 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

$$\begin{aligned}
&p(W^* = w) = p(a\tau(W) + b = w) = p\left(\tau(W) = (w - b)/a\right) \\
&q(W^* = w + a\Delta) = q(a\tau(W) + b = w + a\Delta) = q\left(\tau(W) = (w - b)/a + \Delta\right)
\end{aligned}$$

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 $\Delta$), $w_q^* = w_p^* + a\Delta$, 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 regular and broad-ranging, and consider the generalised hypotheses CIRA$_\tau$, Full-range$_D$ and Normalisation$_{r,s}$ for given $D, \tau, r, s$.

1. In this part, we fix a well-behaved welfare function $W$ satisfying the generalised hypotheses, and we prove that $W$ has the specified form. By Lemma 7, the transformed welfare function

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

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 as given in Theorem 1. Defining $\rho$ 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. Q.e.d..

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

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

with $U$ and $\rho$ defined as in Theorem 2, is well-behaved and satisfies CIRA$_\tau$, Full-range$_D$ and Normalisation$_{r,s}$. By Lemma 7, this is the case if the transformed welfare function $W^*$ defined by (1) is well-behaved and satisfies CIRA, Full-range and Normalisation. By plugging the expression defining $W$ into the one 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 as 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


Abdellaoui, M., Barrios, C., Wakker, P. (2007) Reconciling introspective utility with revealed preference: Experimental arguments based on prospect theory, *Journal of Econometrics* 138(1): 356–378

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

Adler, M. D. (2019) *Measuring Social Welfare: An Introduction*, 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

Baccelli, J. (2024) Ordinal utility differences, *Social Choice and Welfare* 62: 275–287

Baccelli, J., Mongin, P. (2016) Choice-Based Cardinal Utility, *Journal of Economic Methodology* 23(3): 268–288

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

Bernoulli, D. (1738) Specimen Theorize Naval de Mensura Sortis, *Commentarii Academiae Scientiarum Imperialis Petropolitanae* V: 1730–1731 [translated in Sommer, L. (1954) Exposition of a New Theory on the Measurement of Risk, *Econometrica* 22(1): 22–36]

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

Cibinel, P. (2024) Welfare and autonomy under risk, Princeton University, private communication

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., Blanchet, D. (2013) *Beyond GDP: Measuring Welfare and Assessing Sustainability*, Oxford University Press

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

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

Karni, E., Weymark, J. (2024) Impartiality and Relative Utilitarianism, *Social Choice and Welfare* 63(1): 1–18

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

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

Nebel, J. (foth.) Extensive Measurement in Social Choice, *Theoretical Economics*

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