

Interpersonally Comparable Utility

Pete Caradonna¹ Zachary Raines²

¹Division of the Humanities and Social Sciences
California Institute of Technology

²Sustaining Engineering Group
Canonical USA Inc.

On the Measurement of Utility

“If we want to compare utilities between two people much the same problem exists . . . we do not seem to have any ‘outside thing’ which can be measured by both persons to ascertain the relation between the units.”

-Luce & Raiffa (1957)

Introduction

- The question of how to obtain cardinally meaningful, and interpersonally comparable, measures of well-being is among the most foundational in economics.
- However, despite centrality, these notions have frustrated nearly all attempts to give them formal meaning within the standard framework of economic theory.
 - Led to broad consensus that any methodology must necessarily rely on normative, external judgments (e.g. Robbins 1938, Arrow 1950, Sen 1999, Fleurbaey and Hammond 2004).

Objectives

- The objective of this paper is to develop theory of comparable utilities starting from standard, *ordinal* primitives.
- Rather than emphasize particular functional form of representation, we introduce notion of 'measurement unit' for utilities.
- This provides natural notion of comparability: we define a profile of utilities to be interpersonally comparable precisely when every representation is denominated in a common unit.

A Disclaimer

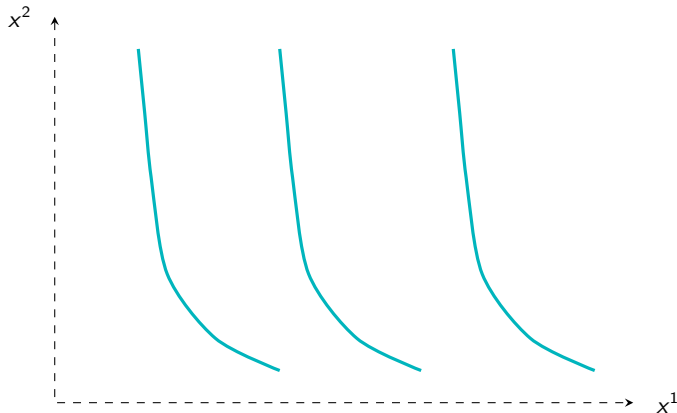
This is **not** a positive theory of utility—we make no claim that comparable profiles reflect literal felicities experienced by individuals.

Instead, our notion provides benchmark for when it is justifiable, or *rationalizable*, for planner to treat utility differences across pairs of both (i) outcomes, and (ii) individuals, as quantified on common scale.

A Simple Example

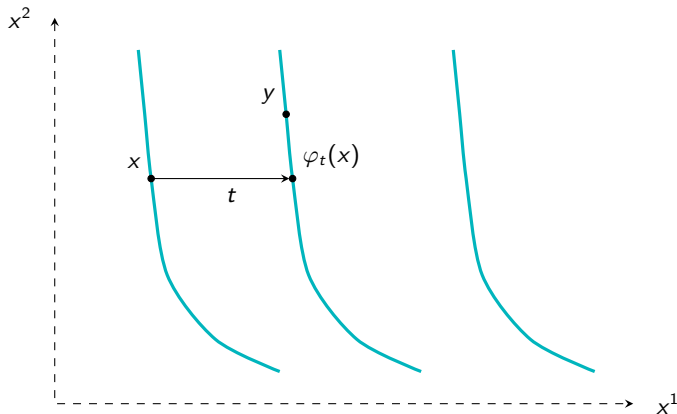
Suppose society consists of S individuals, with preferences over bundles $(x^1, x^2) \gg 0$, that can be represented by quasilinear utility functions:

$$u^s(x^1, x^2) = v^s(x^2) + x^1.$$



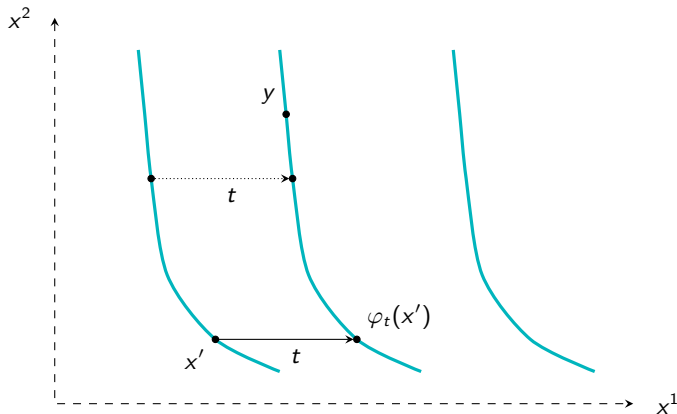
A Simple Example

Let $\varphi_t(x) = x + (t, 0)$. We can use the family $\{\varphi_t\}$ as system of *measuring tapes* to consistently quantify preference intensity:



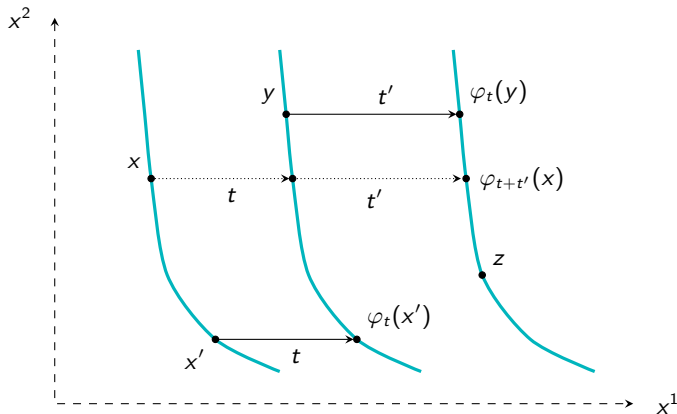
A Simple Example

Let $\varphi_t(x) = x + (t, 0)$. We can use the family $\{\varphi_t\}$ as system of *measuring tapes* to consistently quantify preference intensity:



A Simple Example

Let $\varphi_t(x) = x + (t, 0)$. We can use the family $\{\varphi_t\}$ as system of *measuring tapes* to consistently quantify preference intensity:



Measurement Units Determined by Preferences

This is immediate consequence of quasilinear utilities u^s satisfying:

$$u^s(\varphi_t(x)) = u^s(x) + t. \quad (*)$$

However: the preference $\succsim^s$, not u^s , determines scale unit for numeraire good.

Illustration

- Consider change-of-unit $(\tilde{x}^1, x^2)$, where $\tilde{x}^1 = f(x^1)$, with f nonlinear. Then:

$$u^s(\tilde{x}^1, x^2) = v^s(x^2) + f^{-1}(\tilde{x}^1).$$

- Similarly, can define $\tilde{\varphi}_t(\tilde{x}^1, x^2) = (\tilde{x}^1 + t, x^2)$. But now, $(*)$ doesn't hold, for $\tilde{\varphi}_t$, for *any* increasing transform of u^s .

→ If $(*)$ held, would imply MRS independent of $\tilde{x}^1$, but f is non-linear so $(f^{-1})'$ non-constant!

This Paper

Paper organized around **three basic questions**:

- (i) When can units of one utility function be regarded as the same as those of another?
- (ii) How should a benevolent planner aggregate a profile of utilities that share a common measurement unit?
- (iii) When do there exist commensurable utilities representing a given profile of preferences, and how unique are they?

For Today: Focus primarily on (i) and (ii). Illustration of (iii) for case of EU preferences.

Preliminaries

Let $\mathcal{X}$ be a d -dimensional, relatively open, convex set of social states, or **alternatives**. We assume $d \geq 2$.

The set of individuals comprising society is denoted $\mathcal{S} = \{1, \dots, S\}$.

A **utility function** is any smooth and critical-point-free map $u : \mathcal{X} \rightarrow \mathbb{R}$. Similarly a **preference relation** is a complete and transitive binary relation $\succsim$ such that:

$$x \succsim y \quad \Longleftrightarrow \quad u(x) \geq u(y)$$

for some utility function u .

Toward an Abstract Notion of Commodity

A **commodity domain** $\mathcal{D} \subseteq \mathcal{X} \times \mathbb{R}$ is an open set, such that for every $x \in \mathcal{X}$:

$$\mathcal{D}^{(x)} = \{t : (x, t) \in \mathcal{D}\}$$

is an open interval, containing zero.

A smooth function $\varphi : \mathcal{D} \rightarrow \mathcal{X}$ is **compensation function** if it satisfies:¹

- (i) For all $x \in \mathcal{X}$, $\varphi_0(x) = x$.
- (ii) For all $x \in \mathcal{X}$, if $t \in \mathcal{D}^{(x)}$ and $t' \in \mathcal{D}^{(\varphi_t(x))}$, then $\varphi_{t'+t}(x)$ is well-defined, and equal to $\varphi_{t'}(\varphi_t(x))$.

¹For clarity, we'll write $\varphi_t(x)$ for $\varphi(x, t)$.

Virtual Commodities

Definition

A **virtual commodity** is a pair $(\mathcal{D}, \varphi)$, where $\mathcal{D}$ is a commodity domain and φ a transfer function, which is *maximal*, i.e. there is no pair $(\mathcal{D}', \varphi')$ such that $\mathcal{D} \subsetneq \mathcal{D}'$ and $\varphi'|_{\mathcal{D}} = \varphi$.

Intuition: A virtual commodity defines a rule for augmenting an allocation x with any feasible quantity $t \in \mathcal{D}^{(x)}$ of some abstract, possibly non-linear, composite good.

Examples

Example

Suppose $\mathcal{X} = \mathbb{R}_{++}^d$. Any classical commodity can be represented as a virtual commodity. For any $i = 1, \dots, d$,

$$\varphi_t(x) = (x^1, \dots, x^i + t, \dots, x^d)$$

defines a virtual commodity, whose commodity domain is given by $\mathcal{D}^{(x)} = (-x^i, \infty)$. However, many other types of transformations also yield virtual commodities. For example,

$$\tilde{\varphi}_t(x) = e^t x$$

defines a virtual commodity with $\tilde{\mathcal{D}} = \mathbb{R}_{++}^d \times \mathbb{R}$.

Numeraire Commodities

Definition

A virtual commodity $(\mathcal{D}, \varphi)$ defines a **virtual numeraire** for a preference $\succsim$ if it satisfies the following axioms:

(i) **Invariance:** For all $x, y \in \mathcal{X}$ and all $t \in \mathcal{D}^{(x)} \cap \mathcal{D}^{(y)}$:

$$x \succsim y \iff \varphi_t(x) \succsim \varphi_t(y).$$

(ii) **Monotonicity:** For all $x \in \mathcal{X}$ and $t \in \mathcal{D}^{(x)}$,

$$\varphi_t(x) \underset{(\succsim)}{\sim} x \iff t \underset{(>)}{\geq} 0$$

Measurement of Preference Intensity

Any virtual numeraire gives rise to system of 'measuring sticks' for assigning *cardinal* measure of preference intensity to a preference, just like quasilinear case.

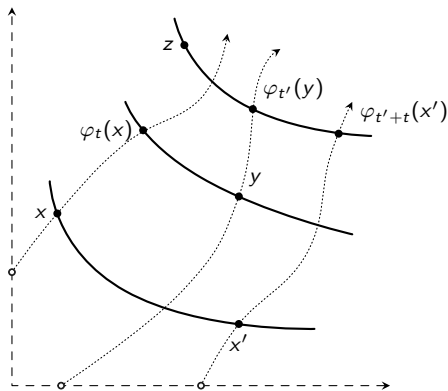


Figure: Here, agent is “willing to pay” t units of numeraire to swap x for y , and t' to swap y for z .

Existence of Measurements

Proposition

Let $(\mathcal{D}, \varphi)$ be a virtual numeraire for $\succsim$. Then for any $x, y \in \mathcal{X}$, there exists a finite sequence $(x_1, t_1), \dots, (x_K, t_K) \in \mathcal{D}$, such that:

$$x \sim x_1, \quad \varphi_{t_1}(x_1) \sim x_2, \quad \dots, \quad \varphi_{t_K}(x_K) \sim y.$$

Moreover, if $(\tilde{x}_1, \tilde{t}_1), \dots, (\tilde{x}_L, \tilde{t}_L) \in \mathcal{D}$ is any other such sequence:

$$\sum_{k=1}^K t_k = \sum_{l=1}^L \tilde{t}_l.$$

Intuition: Even when $\mathcal{D}$ limits how much commodity can be added to any particular alternative, there is always a unique total compensating transfer between any x and y .

Utility Units

Let u be a utility representation of some preference $\succsim$. We say that a virtual numeraire $(\mathcal{D}, \varphi)$ for $\succsim$ is a **utility unit** for u , if for all $x \in \mathcal{X}$ and all $t \in \mathcal{D}^{(x)}$:

$$u(\varphi_t(x)) = u(x) + t.$$

- When $(\mathcal{D}, \varphi)$ is a utility unit for u , its intensity measurements are precisely the utility differences under u .
- However, numeraire willingness-to-pay measurements are well-defined *without any reference to utility*.

Numeraire Measurements are Utility Measurements

Theorem

Let $\succsim$ be a preference on $\mathcal{X}$. Then:

- (i) Every virtual numeraire for $\succsim$ is the utility unit of some choice of representation for $\succsim$.
- (ii) For every representation u of $\succsim$, there exists a virtual numeraire for $\succsim$ which is a utility unit for u .

Moreover, any two representations of $\succsim$ share a common utility unit if, and only if, they differ by an additive constant.

Takeaway: Up to additive constant, utility is uniquely determined by (i) preference, and (ii) measurement unit (i.e. numeraire).

Interpersonally Comparable Utility

Suppose $(u^1, \dots, u^S)$ are a profile of utility functions on $\mathcal{X}$.

Definition

We say $(u^1, \dots, u^S)$ are **interpersonally comparable** if there exists a common utility unit, i.e. a virtual commodity $(\mathcal{D}, \varphi)$ such that, for all $s \in \mathcal{S}$ and $(x, t) \in \mathcal{D}$:

$$u^s(\varphi_t(x)) = u^s(x) + t.$$

Discussion

- We interpret interpersonally comparable as saying that it is justifiable, or *rationalizable*, for planner to treat $(u^1, \dots, u^S)$ as denominated in a common unit.
 - Not just a normalization—generic profiles need not be comparable. In particular, comparability is falsifiable.
 - E.g. no profile containing two representations of the same preference that differ by a non-linear transform can be comparable.
- Requires only that each u^s agree on utility difference across pairs of alternatives of the form x and $\varphi_t(x)$, where $(x, t) \in \mathcal{D}$. Other utility differences unrestricted.
 - In language of Luce & Raiffa, pairs $\{x, \varphi_t(x)\}_{(x,t) \in \mathcal{D}}$ are precisely the ‘outside thing’ which each utility can measure to verify equivalence of units.

An Example

Example

Let $\mathcal{Z}$ be finite set of prizes, partitioned into subsets of good and bad prizes, G and B . Suppose preferences are over $\mathcal{X} = \text{int } \Delta(\mathcal{Z})$, i.e. full-support lotteries on $\mathcal{Z}$, and can be represented by:

$$u^s(p) = \ln \left(\frac{p(G)}{p(B)} \right) + v^s(p|_G, p|_B),$$

where $p|_G, p|_B$ denote the conditional distributions on G and B . Any such profile is comparable, with common unit given by the commodity (with $\mathcal{D} = \mathcal{X} \times \mathbb{R}$):

$$\varphi_t^j(p) = \frac{\exp(t \mathbb{1}_{j \in G}) p^j}{p(B) + \exp(t) p(G)}, \quad j \in \mathcal{Z}.$$

Here, ‘adding’ more commodity shifts mass from B to G , while preserving all the conditional probabilities $p|_G$ and $p|_B$.

A Simple Characterization

Let $r^S(x)$ denote the rank of the $d \times S$ matrix given by the gradients of $(u^1, \dots, u^S)$ at $x \in \mathcal{X}$. Say a profile has **constant rank** if $r^S(x)$ is constant.

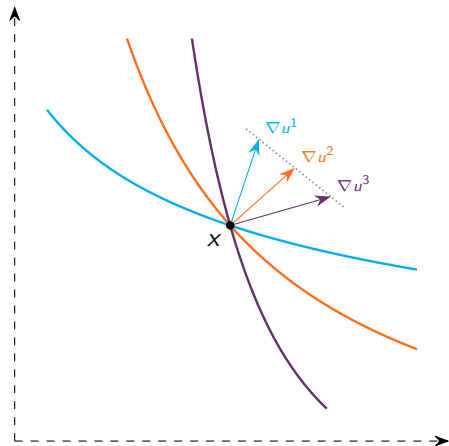
Theorem

Suppose $(u^1, \dots, u^S)$ has constant rank. Then it is interpersonally comparable if, and only if, for all $x \in \mathcal{X}$:

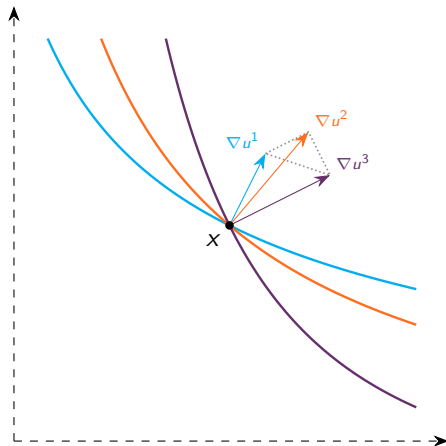
$$0 \notin \text{Aff}\left\{u^1(x), \dots, u^S(x)\right\}.$$

In The Paper: Alternatively, a profile is comparable, with common unit $(\mathcal{D}, \varphi)$, if and only if there is a smooth change of coordinates that renders every u^s simultaneously quasilinear, and transforms φ into translations along an axis. [► Details](#)

An Illustration



(a) Affine hull of gradients at x don't contain 0, consistent with a common unit.



(b) Gradients at x span $\mathbb{R}^2$ hence no common utility unit can exist.

Comparability is Cardinal

When $(u^1, \dots, u^S)$ is comparable, even if an identical, monotone transform is applied to each utility, the resulting profile will generally no longer be comparable.

Proposition

Suppose $S \geq 2$, and let $f = (f^1, \dots, f^S)$ be a tuple of strictly increasing monotone transforms. Then f has the property that $(f^1 \circ u^1, \dots, f^S \circ u^S)$ are comparable whenever $(u^1, \dots, u^S)$ are, if, and only if:

$$f^s(t) = \alpha t + \beta^s,$$

for some $\alpha > 0$, $\beta^s \in \mathbb{R}$.

Social Choice for Comparable Profiles

Question: How should a planner synthesize a comparable profile into an aggregate social ranking?

Following Sen (1970), let $\mathcal{U}$ denote all profiles of utilities on $\mathcal{X}$, and $\mathcal{C} \subsetneq \mathcal{U}$ the subset of comparable profiles. A **social welfare functional** is a map $F : \mathcal{U} \rightarrow \mathcal{P}$, the set of all preference relations on $\mathcal{X}$.

- When F is understood, we use $\succeq_u$ to denote the social preference $F(u)$.

Our interest will instead be in **restricted social welfare functionals** $F : \mathcal{C} \rightarrow \mathcal{P}$.

Common-Scale Comparisons

- For any comparable profile $(u^1, \dots, u^S)$, it is meaningful for the planner to make comparisons of the form:

$$u^s(y) - u^s(x) \geq u^{s'}(y') - u^{s'}(x')$$

for any $s, s' \in \mathcal{S}$ and $x, x', y, y' \in \mathcal{X}$.

→ Each side of inequality is measured in some common unit φ .

- By Bossert (1991), comparisons of this form are *exactly* those preserved by transformations:

$$(u^1, \dots, u^S) \mapsto (\alpha u^1 + \beta^1, \dots, \alpha u^S + \beta^S),$$

where $\alpha > 0$ and each $\beta^s \in \mathbb{R}$.

Unit Comparability

To encode the ability of the planner to make use of precisely this information, we impose the following axiom of Sen (1970).

Definition

A restricted social welfare functional satisfies **cardinal unit-comparability (CUC)** if, for all $u \in \mathcal{C}$:

$$F(u^1, \dots, u^S) = F(\alpha u^1 + \beta^1, \dots, \alpha u^S + \beta^S)$$

for every $\alpha > 0$ and $\beta \in \mathbb{R}^S$.

Note: This is well-defined as $\mathcal{C}$ is closed under such transformations.

Arrowian Social Choice Axioms

The remainder of our axioms are simply the (restricted) social welfare functional axioms of Arrow (1951). Our treatment here follows d'Aspremont & Gevers (1977).

- (IIA) **Independence of Irrelevant Alternatives:** For all $x, y \in \mathcal{X}$ and $u, v \in \mathcal{C}$, if $u(x) = v(x)$ and $u(y) = v(y)$ then $x \succeq_u y$ if and only if $x \succeq_v y$.
- (SP) **Strong Pareto:** For all $u \in \mathcal{C}$ and $x, y \in \mathcal{X}$, $u^s(y) \geq u^s(x)$ for every $s \in \mathcal{S}$ implies $y \succeq_u x$. If, in addition, $u^s(y) > u^s(x)$ for some $s \in \mathcal{S}$, then $y \succ_u x$.
- (A) **Anonymity:** For any permutation σ of $\mathcal{S}$ and any $u \in \mathcal{C}$, $F(u^1, \dots, u^S) = F(u^{\sigma(1)}, \dots, u^{\sigma(S)})$.

Utilitarianism of Comparable Profiles

Theorem

Suppose $F : \mathcal{C} \rightarrow \mathcal{P}$ is a restricted social choice functional for comparable profiles. Then F satisfies (CUC), (IIA), (SP), and (A) if and only if:

$$x \succeq_u y \iff \sum_{s \in \mathcal{S}} u^s(x) \geq \sum_{s \in \mathcal{S}} u^s(y)$$

for all $x, y \in \mathcal{X}$ and $u \in \mathcal{C}$.

Takeaway: An Arrovian social planner, faced with a comparable profile of utilities, is utilitarian.

Discussion

- d'Aspremont & Gevers (1977) prove analogous result for unrestricted social welfare functionals. Our result differs methodologically in a few respects:
 - We restrict attention to only comparable profiles. Proof requires new richness type arguments to compensate.
 - More structured notion of preferences/utility and richer domain.
- Relative to social choice literature, novelty is in our use of comparability to justify assumption of unit-comparability, rather than imposing it exogenously.
 - See, e.g., Fleurbaey & Hammond (2004).

Ordinal Primitives

Suppose now a planner observes only a *preference* profile $(\succsim^1, \dots, \succsim^S)$.

- (i) Given choice of common virtual numeraire $(\mathcal{D}, \varphi)$, there exists a unique (up to individual additive constants) comparable profile of representations $(u^1, \dots, u^S)$.
 - (ii) An Arrovian planner will then seek to maximize sum of these utilities (in particular, individual additive constants have no effect).
- Thus, each common virtual numeraire gives rise to equally justifiable social preference.

Question: What is the *set* of common virtual numeraires for a given preference profile?

In The Paper

- Complete characterization in the case of local welfare analysis, as well as several global results.
- Methodologically, we rely on set of differential geometric techniques due to Cartan (1908). Widespread applications in physics, geometry, optimization and control theory.
- For sufficiently rich profiles, set of common virtual numeraire depends smoothly on at most $\dim \mathcal{X} + 1$ parameters.
 - For sufficiently rich preference profiles, set of comparable tuples of utilities is finite-dimensional—a vanishingly small proportion.
 - Often take on recognizable form.

Application: Expected Utility Preferences

Suppose $\mathcal{Z} \subset \mathbb{R}$ is a finite set, and $\mathcal{X} = \text{int } \Delta(\mathcal{Z})$. Let $(\succsim^1, \dots, \succsim^S)$ be a profile of strictly monotone EU preferences. Several well-known approaches to aggregation:

- (i) **EU Utilitarianism:** (Harsanyi 1953, 1955) Maximize weighted sum of expected utilities.
- (ii) **Relative Utilitarianism:** (Dhillon & Mertens 1999): Maximize sum of unique expected utility representations, normalized to have range $[0, 1]$.
- (iii) **Nash Social Welfare Function:** (Kaneko & Nakamura 1979) Maximize Nash product of expected utilities, normalized so worst prize gets utility zero.

Application: Expected Utility

Say a virtual commodity is **translationalary** if it is of the form:

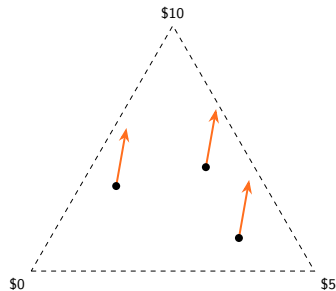
$$\varphi_t(p) = p + t\gamma.$$

When (MON) holds, any such commodity is a common numeraire, with comparable utilities:

$$u^s(p) = \frac{1}{\gamma \cdot \nabla \tilde{u}^s} \tilde{u}^s(p),$$

where $\tilde{u}^s$ is any expected utility representation.

→ For translationalary virtual numeraires, Arrowian planner maximizes *weighted sum of expected utilities*.

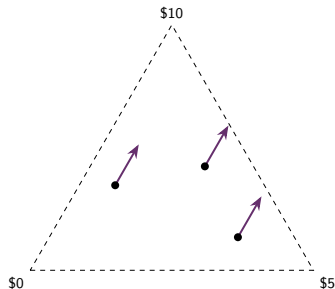


Application: Expected Utility

In special case when $\gamma \propto \delta_{\text{MAX}} - \delta_{\text{MIN}}$, comparable representations remain expected utilities.

Here, normalizations ensure the range of each comparable (expected utility) representation is open interval of common length.

- When γ proportional to $\delta_{\text{MAX}} - \delta_{\text{MIN}}$,
Arrovian planner maximizes increasing affine
transform of *relative utilitarian* objective.



Application: Expected Utility

Conversely, say a virtual commodity is **dilationary** if it is of the form:

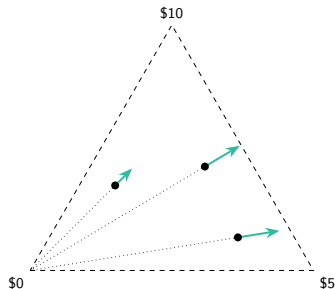
$$\varphi_t(p) = e^{at}p + (1 - e^{at})\delta_{\$0}$$

for some $a > 0$. This always defines a common virtual numeraire. Comparable utilities are now:

$$u^s(p) = \ln(\tilde{u}^s(p) - \tilde{u}^s(\delta_{\$0})),$$

where $\tilde{u}^s$ is any expected utility representation.

- For any dilationary virtual numeraire, an Arrovian planner maximizes the (log of) the *Nash social welfare functional*.



Takeaway

We find (i) utilitarianism of expected utilities, (ii) relative utilitarianism, and (iii) the Nash social welfare functional, all arise as special cases of general theory: *utilitarianism of comparable profiles*.

- While each of these SWFs have existing axiomatizations, we provide unified perspective.
- Clarifies where and how 'normative' choices, in form of selection of implicit measurement unit, enter.

Expected Utility Units

Natural Question: Are there other common numerares, yielding new methods of aggregation?

Theorem

For any sufficiently rich profile of monotone, EU preferences, every common virtual numeraire can be written as a dilation, followed by a translation.

► Richness Details

Why Specifically Expected Utilities?

A longstanding critique of these methods of aggregation is that maximizing specifically the sum (or product) of *expected utilities* is arbitrary and unjustified:

- *“This theorem does not, as far as I can see, give any special ethical significance to the particular utility scale found. For instead of using the utility scale found by von Neumann and Morgenstern, we could use the square...” Arrow (1951)*
- *“[F]rom a cardinal point of view, the Nash social welfare seems to be equally motivated under arbitrary monotone transformations.” Gonczarowski & Segev (2026)*
- *“What justifications have been offered for singling out von Neumann–Morgenstern utility functions from the class of all representations of $\succeq$? The main justification has been that of convenience.” Weymark (2005)*

A New Perspective

Our theory clarifies expected utilities (and their logs) occupy privileged position among all possible representations: they compose the *comparable profiles*.

- Obtains endogenously: none of our theory makes any reference to risk.

Relates to longstanding debate between Sen, Roemer, Weymark et al. versus Harsanyi over whether Harsanyi's theorems provide foundations for utilitarianism.

- Debate focused on whether expected utilities are interpersonally comparable.
- Historically, this meant whether the vNM representation supplies *positive* theory of comparable utility, rather than our 'as-if' sense.
 - **Our Results:** if *any* utilities are interpersonally comparable, necessarily they are either EU or normalized logarithm of EUs.

Conclusions

- Problem of obtaining interpersonally-comparable cardinal utilities is of central importance to much of economics.
- However, despite importance, these concepts have frustrated nearly all attempts to give them formal meaning within standard framework of economic theory.
- This paper develops theory of measurement units for utility, rich enough to supply appropriate units for any representation, of any preference.
- From this, we obtain natural, transparent criterion for comparability of utilities. Remainder of paper explores implications of this definition.

Thank You!

Any Questions?

Quasilinearity as a Normal Form

Under a mild regularity condition on $(\mathcal{D}, \varphi)$... [► Details](#)

Theorem

A virtual commodity $(\mathcal{D}, \varphi)$ is a common utility unit for a profile $(u^1, \dots, u^S)$ if, and only if, there exists a smooth manifold $\mathcal{Y}$, open subset $\Omega \subseteq \mathbb{R} \times \mathcal{Y}$, and smooth change-of-coordinates $h : \mathcal{X} \rightarrow \Omega$ such that:

(i) For all $(x, t) \in \mathcal{D}$,

$$h(\varphi_t(x)) = h(x) + (t, 0, \dots, 0).$$

(ii) For all $s \in \mathcal{S}$ and $(t, y) \in \Omega$:

$$u^s(h^{-1}(t, y)) = t + v^s(y),$$

for some smooth function v^s .

Regular Virtual Commodities

Definition

Suppose $(\mathcal{D}, \varphi)$ is a virtual commodity. Define the equivalence relation:

$$x \sim_{\varphi} y \iff y = \varphi_t(x) \text{ for some } t \in D^{(x)}, \text{ or vice-versa.}$$

We say that $(\mathcal{D}, \varphi)$ is a **regular** virtual commodity if $\sim_{\varphi}$ is a continuous equivalence relation on $\mathcal{X}$, i.e. its graph is closed in $\mathcal{X} \times \mathcal{X}$. [▶ Return](#)

Definition

Say that a profile $(\succsim^1, \dots, \succsim^S)$ of monotone EU preferences is **rich** if it contains a sub-profile $(\succsim^{s_1}, \dots, \succsim^{s_{|Z|}})$, such that:

- (i) $\nabla \tilde{u}^{s_1}, \dots, \nabla \tilde{u}^{s_{|Z|-1}}$ are (everywhere) linearly independent; and
- (ii) $\nabla \tilde{u}^{s_{|Z|}}$ is a linear combination of the $\nabla \tilde{u}^{s_1}, \dots, \nabla \tilde{u}^{s_{|Z|-1}}$, with non-zero coefficients.