XU

Loren K. Fryxell

University of Oxford and Global Priorities Institute

loren.fryxell@economics.ox.ac.uk

June 17, 2024

Oxford Research Jamboree 2024

Introduction

Overview

This paper does two things.

1. It presents a theory of ratio-scale utility. I call this XU, which stands for *experienced utility*.
2. It presents a principle of ethical choice. I call this the LELO principle, which stands for *live every life once*.

XU

XU

XU is a theory of ratio-scale utility which captures an individual's intensity of preference

Two things to note

1. ratio-scale utility
2. intensity of preference

Ratio-Scale Utility

Ratio-scale utility means that

1. **Meaningful Zero Point.** Zero utility is a meaningful concept. Adding a constant to the utility will meaningfully change the underlying primitive.
2. **Meaningful Intervals.** Comparisons of utility differences is meaningful. Changing the difference in utility between any pair of alternatives, relative to other pairs, will meaningfully change the underlying primitive.
3. **Meaningful Ratios.** The ratio of utilities is meaningful. Changing the ratio between a pair of alternatives will meaningfully change the underlying primitive. For instance, it is meaningful to say that one alternative has twice as much utility as another.

Preference Intensity

Preference intensity captures *how much* you like something, like an apple

Intuitively, suppose I say I like an apple 100x more than a cantaloupe,

and I like a banana 2x more than a cantaloupe

Presumably, there is something important captured here more than just

$$\text{apple} \succ \text{banana} \succ \text{cantaloupe}$$

Preference Intensity vs Risk Attitudes

What about preferences over gambles?

Usually in economics, we use preferences over gambles (vNM expected utility) to get at this notion of preference intensity

But this gets at an individual's **risk attitudes**, not their **preference intensity**

An individual's risk attitude is conceptually distinct from "how much more" they like an apple than a banana

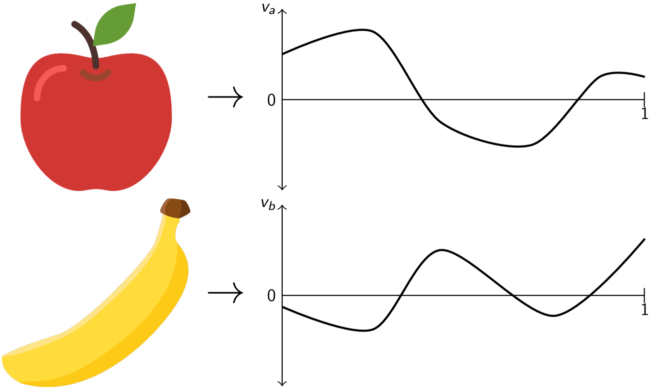
Luce and Raiffa (1957) identify conflating risk attitudes and preference intensity as "Fallacy 3"

What is XU?

XU measures utility as it is experienced **over time**

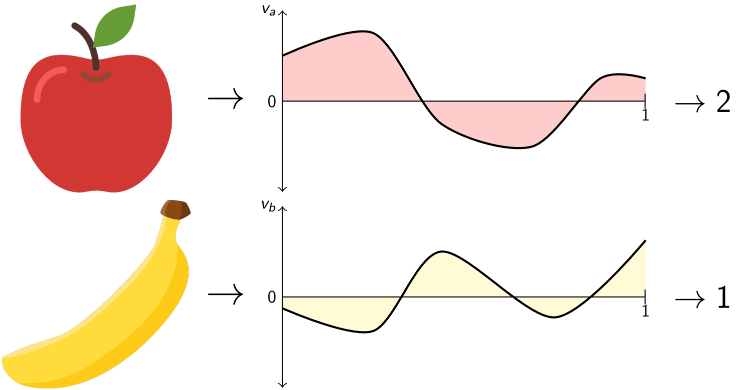
XU

For each alternative, we construct a measure of instantaneous preference intensity over time (a hedonometer)



XU

The area under this curve provides a ratio-scale utility value for the alternative, capturing how good it is for the individual



LELO

LELO

LELO stands for *live every life once*.

The LELO Principle. *The choice that is best for society is the choice that an ethical observer would most prefer if they were to live every life once.*

The LELO Principle vs The Pareto Principle

LELO can be thought of as a much stronger guiding principle than the Pareto principle

The Pareto principle captures a notion of “obviously better than”. It is easy to agree with and requires little information but it rarely has anything to say (it is an incomplete ordering)

The LELO principle is a much stronger principle. It takes more to agree with, requires more information, but it has something to say every time (it is a complete ordering)

Putting XU and LELO Together

XU and LELO

Result. An ethical observer who respects LELO acts *as if* they are maximizing total XU across individuals.

That is, LELO implies utilitarianism with respect to some utility representation and that representation is XU.

The Hedonometer

Main Idea

The main idea of this project is that any choice set is, in essence, a set of **experiences** (something that occurs over time)

When you say "apple $\succeq$ banana", you mean

what will happen if you are given the apple
$\succeq$ *what will happen if you are given the banana*

Main Idea

If we consider a choice set as a set of **experiences** that occur over time, there are many more preference/choice comparisons we can make over simply

apple vs banana

For example,

first minute of eating apple vs first minute of eating banana

Only considering preferences over "complete experiences" (entire apple vs entire banana) is leaving a lot of information on the table

Considering preferences over all time slices will allow us to construct **ratio-scale utilities** that capture an individual's **intensity of preference**

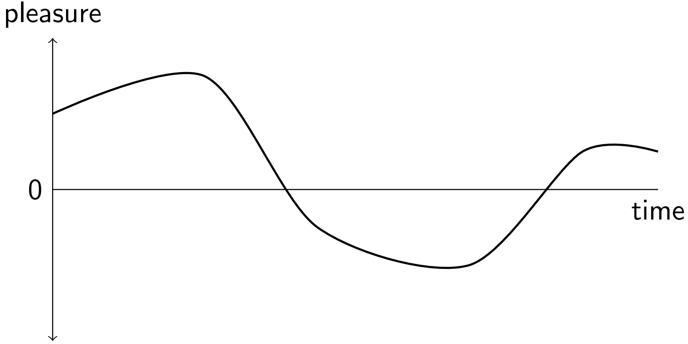
Experienced Utility

There is a long history of thinking about utility as a measure of pleasure over time

Goes all the way back to Edgeworth's "hedonometer"

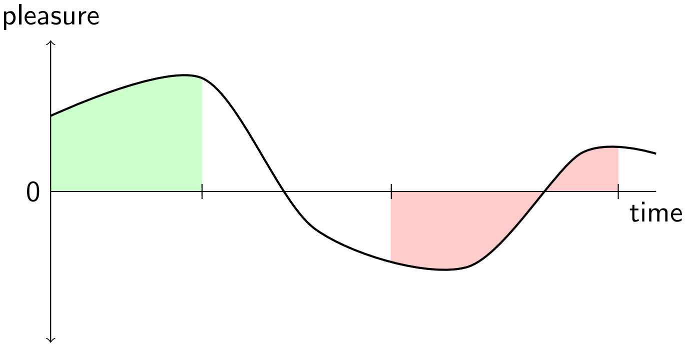
Edgeworth's Hedonometer

Imagine an ideally perfect instrument, a psychophysical machine, continually registering the height of pleasure experienced by an individual.



Edgeworth's Hedonometer

The quantity of happiness between two epochs is represented by the area contained between the zero-line ... and the curve traced by the index.

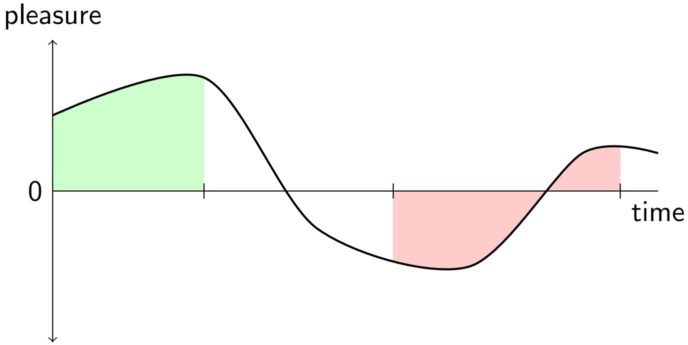


The green time slice gives more "happiness" than the red time slice

XU provides an axiomatic preference-based foundation for such a notion of experienced utility over time

Main Idea

The main idea is simple. A hedonometer should represent an individual's *preferences* over living these experiences



The individual prefers to live the green time slice to the red time slice

Axioms

I propose axioms on $\succeq$ over experiences that are necessary and sufficient for such a representation (a hedonometer) to exist

Measurement Theory

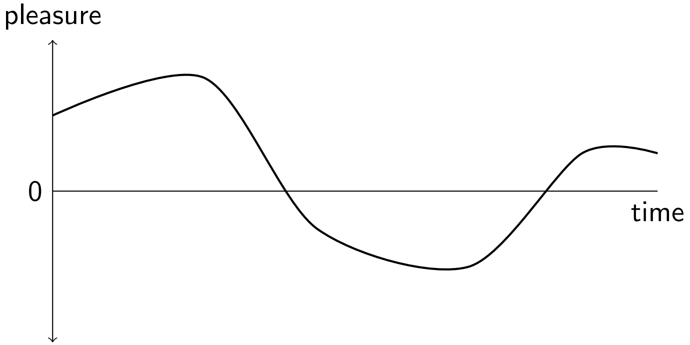
Zooming out, the formal result is a more general contribution to *measurement theory*

In particular, we need not think of the x-axis as time, the y-axis as pleasure, and the relation $\succeq$ as a preference

(Ask me after if you're interested)

XU

What is this timeline?



I will call a timeline such as this an *event*

An **event** is a complete description of everything that is happening at each point in time within a time slice (normalized to the unit interval)

Single Event

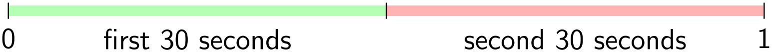
In this section, we will have just a single event. To fix ideas, consider

eating an apple over one minute

This includes a description of the individual's *state of mind* at each point in time, including how full she feels

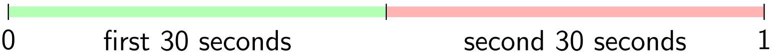
Experiences

An **experience** is any sub-interval within an event (including the entire event)



Preferences over Experiences

We will suppose that an individual has preferences over experiences



$[0, .5) \succ [.5, 1)$ means that the individual prefers

- the experience of the first 30 seconds of eating an apple
- to the experience of the second 30 seconds of eating an apple

For example, she enjoys the apple more when she is hungry, and she feels more full during the second 30 seconds of eating an apple than she does during the first

Interpretation of $\succeq$

Thinking carefully about what it means to say $I \succ J$, or equivalently what it means to choose I from $\{I, J\}$, is a crucial part of the philosophical underpinnings of this project

Indeed, I could probably spend an entire presentation just on this

But given time constraints, all I will say now is the following

Interpretation of $\succeq$

1. How you define what it means to say $I \succ J$ completely changes the interpretation of the axioms (and if they are sensible) and the meaning resulting representation
2. There are several useful and insightful ways to define what this means
   - (indeed one extreme example I won't discuss is to interpret this as a notion of "heavier than", in which case the representation is a measure of weight)
3. However, I propose **one in particular** for which I argue the resulting representation is a sensible measure of an individual's **well-being**

Interpretation of $\succeq$

I want you to think of $I \succ J$ to mean that you would rather *live* the experience of I rather than to live the experience of J

Formal Definitions

An experience is represented formally by an **interval**

$\mathcal{I}$ is the set of all left-closed, right-open, **non-empty** intervals that lie in $[0, 1]$

Left-closed, right-open so adjacent intervals fit together nicely:

- adjacent intervals are disjoint
- the union of adjacent intervals is itself an interval

If I and J are adjacent, then define concatenation by $I \oplus J \equiv I \cup J$

Notice that $\mathcal{I}$ is closed under $\oplus$. For example, $[0, .2) \oplus [.2, .5) = [0, .5) \in \mathcal{I}$

Axioms

There are three axioms:

1. Rationality
2. Monotonicity
3. Continuity

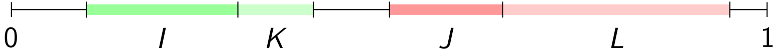
Axiom 1. Rationality

Axiom 1. An individual has preferences $\succeq$ over $\mathcal{I}$.

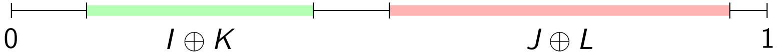
Axiom 2. Monotonicity

Suppose I and K are adjacent, and J and L are adjacent

If $I \succeq J$ and $K \succeq L$



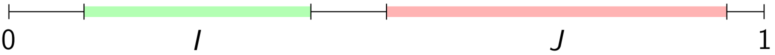
then $I \oplus K \succeq J \oplus L$



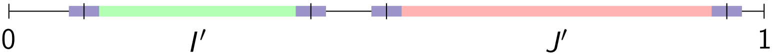
Axiom 3. Continuity

Small changes in an interval result in small changes in preference

If $I \succ J$



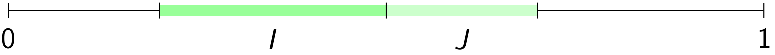
then $I' \succ J'$



Theorem 1 Preliminaries

Before stating the results, we must first define the concept of a positive, zero, and negative experience

Positive, Zero, and Negative Experiences



Definition. An incomplete experience $I \in \mathcal{I} \setminus [0, 1)$ is

1. *positive* if there exists an experience $J \in \mathcal{I}$ such that $I \oplus J \succ J$, i.e. adding I to J makes it better,
2. *zero* if there exists an experience $J \in \mathcal{I}$ such that $I \oplus J \sim J$, i.e. adding I to J makes no difference, and
3. *negative* if there exists an experience $J \in \mathcal{I}$ such that $I \oplus J \prec J$, i.e. adding I to J makes it worse.

Aside

As defined, an experience could, in principle, be all three at once

Proposition. Suppose Axioms 1 and 2.

1. For any $I \in \mathcal{I}$, if $I \oplus J \succ(\sim)(\prec)\ J$ for some adjacent J, then $I \oplus J' \succ(\sim)(\prec)\ J'$ for all adjacent J'.
2. Every incomplete experience is exactly one of positive, zero, and negative.

Positive, Zero, and Negative Experiences

We still don't have a definition of a positive, zero, and negative *complete* experience (the entire unit interval)

Definition. The complete experience $[0, 1)$ is

1. *positive* if there exists an incomplete positive experience $P \in \mathcal{I} \setminus [0, 1)$ such that $[0, 1) \succeq P$,
2. *zero* if there exists an incomplete zero experience $Z \in \mathcal{I} \setminus [0, 1)$ such that $[0, 1) \sim Z$, and
3. *negative* if there exists an incomplete negative experience $N \in \mathcal{I} \setminus [0, 1)$ such that $[0, 1) \preceq N$.

Aside

As before, a complete experience could, in principle, be all three at once (or in fact none of them)

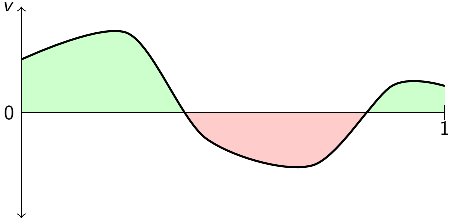
Proposition. Suppose Axioms 1, 2, and 3. The complete experience $[0, 1)$ is exactly one of positive, zero, and negative.

Uniformly Non-Negative and Non-Positive Experiences

Let $\mathcal{I}(I) = \{J \in \mathcal{I} : J \subseteq I\}$ be the set of experiences contained within I

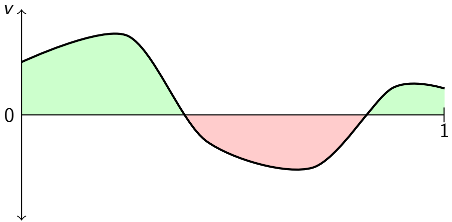
Definition. An experience $I \in \mathcal{I}$ is

1. *uniformly non-negative* if all sub-experiences $J \in \mathcal{I}(I)$ are positive or zero, and
2. *uniformly non-positive* if all sub-experiences $J \in \mathcal{I}(I)$ are negative or zero.



Ordinary

Definition. $[0, 1)$ is *ordinary* if every $t \in [0, 1]$ is either in the interior or on the boundary of some uniformly non-negative or uniformly non-positive segment.



Additive and Continuous Representation

Definition. $U : \mathcal{I} \to \mathbb{R}$ is *additive* if for any $I, J \in \mathcal{I}$,

$$U(I \oplus J) = U(I) + U(J).$$

Definition. $U : \mathcal{I} \to \mathbb{R}$ is *continuous* if for any $I \in \mathcal{I}$ and $\delta > 0$, there exists $\varepsilon > 0$ such that if $I' \in \mathcal{I}$ is less than ε-distant from I, then

$$\left| U(I') - U(I) \right| < \delta.$$

Theorem 1

Theorem 1. Suppose $[0, 1)$ is ordinary. Axioms 1, 2, and 3 hold if and only if there exists an additive and continuous representation $U : \mathcal{I} \to \mathbb{R}$ of $\succeq$. Moreover, $\hat{U}$ is an additive and continuous representation of $\succeq$ if and only if $\hat{U} = \alpha U$ for some $\alpha > 0$ (U is a ratio-scale).

Axiom 3'.

What about an integral?

If we strengthen the continuity axiom, then we can get an integral representation (which is also additive and continuous)

Axiom 3' (Intuitively). Small segments are small in preference.

Theorem 2

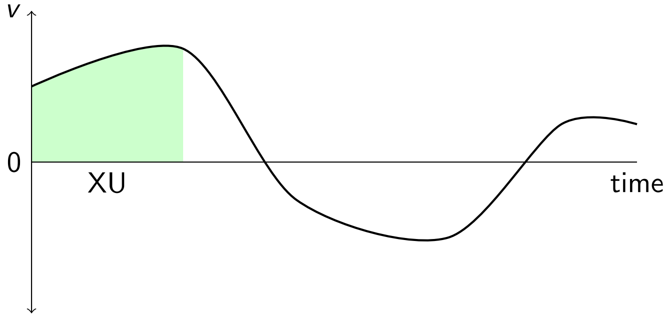
Theorem 2. Suppose $[0,1)$ is ordinary. Axioms 1, 2, and 3' hold if and only if there exists an integral representation $U : \mathcal{I} \to \mathbb{R}$ of $\succeq$, where

$$U(I) = \int_{I^0}^{I^1} v(t)\,\mathrm{d}t$$

and v is Lebesgue integrable. Moreover, $\hat{U}$ is an integral representation of $\succeq$ if and only if $\hat{U} = \alpha U$ for some $\alpha > 0$ (U is a ratio-scale).

The Hedonometer

And so we have our hedonometer



Why an Additive Representation?

Why do we want an additive (or integral) representation?

A numerical representation/measurement is simply a language to talk about the underlying primitive

The underlying primitive in the weight case is how objects will balance on a scale

The underlying primitive in the experienced utility case is how we rank experiences

Why an Additive Representation?

An additive representation is **useful** because it is **intuitive**

To see why this is so, let's consider an alternative representation—a multiplicative one

Let U be an additive and continuous representation of $\succeq$, so that

$$U(I \oplus J) = U(I) + U(J).$$

Let $V = \exp(U)$. Then

$$V(I \oplus J) = V(I)V(J).$$

That is, V is *multiplicative*

Why an Additive Representation?

Let U be additive and $V = \exp(U)$ be multiplicative

- If I is zero, i.e. for any adjacent J, $I \oplus J \sim J$, then
  - $U(I) = 0$
  - $V(I) = 1$
- If I is positive, i.e. for any adjacent J, $I \oplus J \succ (\prec)\, J$, then
  - $U(I) > 0$
  - $V(I) > 1$
- If I_1 and I_2 are two halves of I by preference, i.e. $I_1 \sim I_2$ and $I_1 \oplus I_2 = I$, then
  - $U(I_1) + U(I_2) = U(I)$, and $U(I_1) = U(I_2) = \frac{1}{2}U(I)$
  - $V(I_1)V(I_2) = V(I)$, and $V(I_1) = V(I_2) = \sqrt{V(I)}$
- If I is the worst experience you can imagine, then
  - $U(I)$ is very negative
  - $V(I)$ is positive and close to zero

Why an Additive Representation?

We could easily measure our well-being through V, and there would be nothing incorrect about doing so per se

As long as we define how the representation/measurement maps to the primitive, it is all equivalent

But this would arguably be quite counter-intuitive, and hence not so **useful**

Hence why we seek an additive representation. It is the measure that captures our **intuition**.

Why an Additive Representation?

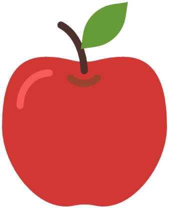
There is nothing magical about adding well-being

What is important is that, whatever representation we use, we understand precisely what it conveys about the primitive

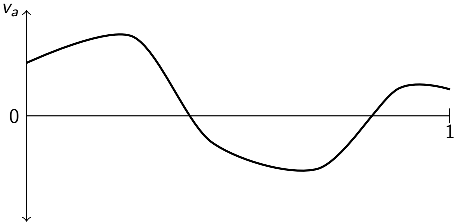
Multiple Events

What We've Done So Far

So far, we have constructed an experienced utility representation for a single event

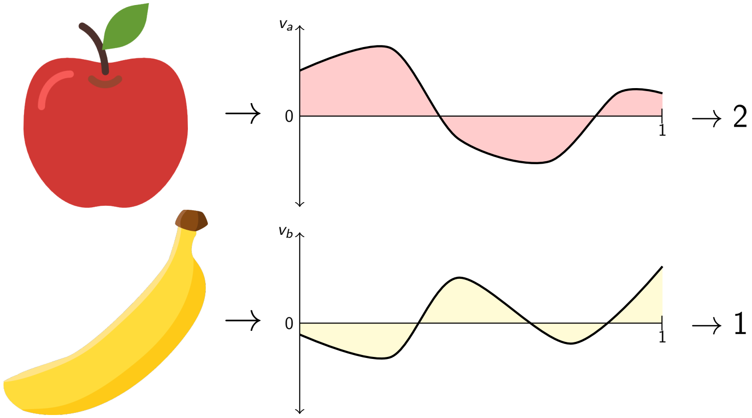


→



Many Events

Now we consider preferences over experiences with an arbitrary set of events X, giving us **ratio-scale utilities** on X



Experiences and Representation

Previously, an *experience* was just an interval

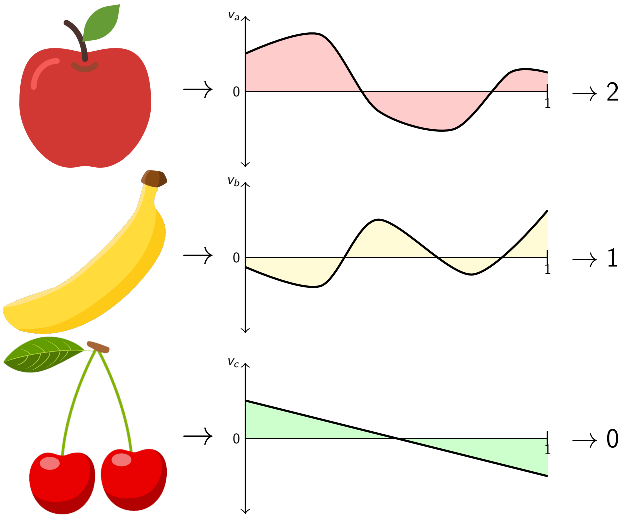
Now, an *experience* is a pair (x, I)

- x is an element of an arbitrary set of events X
- I is an interval

We would like to construct an instantaneous utility function v_x for each event $x \in X$

The axioms and theorem are very similar (we won't go through them)

This Gives Us Ratio-Scale Utilities on X

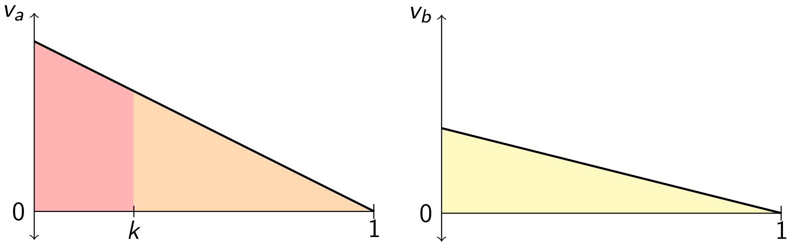


Preference Intensity

So, what does it mean to like an apple twice as much as a banana?

Preference Intensity

An apple is twice as good as a banana **means** that “half” the experience of an apple, by preference, is indifferent to the complete experience of a banana

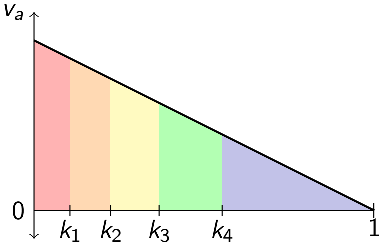


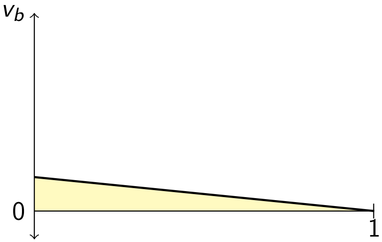
$$(a, [0, k)) \sim (a, [k, 1)) \sim (b, [0, 1))$$

Preference Intensity

We quantify how much better an apple is than a banana by asking:

Into how many equal pieces can we cut the experience of an apple so that each piece is indifferent to the entire experience of a banana?





$$(a, [0, k_1)) \sim \ldots \sim (a, [k_4, 1)) \sim (b, [0, 1))$$

LELO

Extended Sympathy

Recall that an experience is a complete description of everything that is happening over some interval

This includes a description of the state of the mind of the individual experiencing it

But a description of the state of mind includes a description of *who* you are

"Hungry-Ann eating an apple" and "Full-Ann eating an apple" are different experiences

"Hungry-Ann eating an apple" and "Hungry-Bob eating an apple" are also different experiences

Extended Sympathy

Recall that an *experience* is a pair (x, I)

- x is an element of an arbitrary set of events X
- I is an interval

Now, replace X with $X \times N$, where

- X is a set of events
- N is a set of individuals

$(x, i, [0, 1)$ is the experience of event x through the eyes of individual i for the duration $[0, 1)$

Quote

"Ethics is a theory of rational behavior in the service of the common interests of society as a whole" (Harsanyi, 1977).

The LELO Principle

Observation. The outcome that is best for you is the outcome you would prefer if you were to live only as you

Proposal. The outcome that is best for society is the outcome you would prefer if you were to *live every life once* (LELO)

The LELO (Live Every Life Once) Principle.

$$x \succeq^* y \iff \oplus_{i \in N} (x, i, [0,1) \succeq \oplus_{i \in N} (y, i, [0,1)).$$

XU and LELO

Let U be some representation of your preferences $\succeq$ over extended experiences $A \times N \times \mathcal{I}$

Let $U_i(x) = U(x, i, [0,1))$ be the utility of the event x in individual i's shoes

Let $W : \mathbb{R}^N \to \mathbb{R}$ and suppose W is a representation of $\succeq^*$, i.e.

$$x \succeq^* y \iff W(U_1(x), \ldots, U_N(x)) \geq W(U_1(y), \ldots, U_N(y))$$

Proposition. Suppose social preferences $\succeq^*$ satisfies the LELO principle. Then

$$W \text{ is utilitarian} \iff U \text{ is an XU representation.}$$

XU and LELO

Proposition. Suppose social preferences $\succeq^*$ satisfies the LELO principle. Then

$$W \text{ is utilitarian} \iff U \text{ is an XU representation.}$$

What does this mean?

1. If you have an XU representation, you want to add the utilities (be utilitarian)
2. If you have anything other than an XU representation, you **do not want to add the utilities** (**don't** be utilitarian)

A Remark

In my view, utilitarianism is an under-specified theory

I can't decide if I want to add utilities, or do something else, until I know precisely what those utilities are

For example, if you accept the LELO principle, are you a utilitarian?

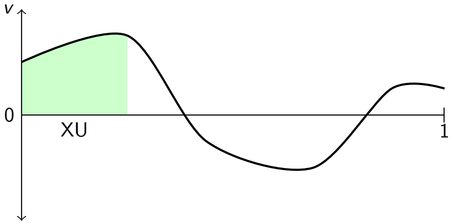
It depends. If XU is your theory of well-being, then maximizing the sum of well-being corresponds to LELO, so yes

If anything other than XU is your theory of well-being, then maximizing the sum of well-being does not correspond to LELO, so no

Conclusion

Conclusion

XU. Preferences over experiences gives rise to ratio-scale utility



LELO. The choice that is best for society is the choice that an ethical observer would most prefer if they were to live every life once.

LELO implies utilitarianism with respect to some utility representation and that representation is XU

Global Priorities Research

Why is this global priorities research?

Global priorities research is research which intends to make progress towards answering crucial questions that arise when trying to do the most good.

In this paper,

1. I propose a framework for thinking about experienced utility (XU), and in particular, a theory of measurement of well-being. This is a theory of what is good for an individual.
2. I present a principle of ethical choice (LELO). This is a theory of what is good for a society.

Global Priorities Research

These theories of the good for the individual (XU) and society (LELO) immediately extend to two important domains not specifically discussed here

- animal welfare (the question of how to value animal lives)
- population ethics (the question of how to value future human or animal lives)

Thank you for listening!

Questions, comments, or concerns?