Foundations of Weighted Utilitarianism*

Rossella Argenziano[†] and Itzhak Gilboa[‡]

October 2017

Abstract

We provide a microfoundation for a weighted utilitarian social welfare function that reflects common moral intuitions about interpersonal comparisons of utilities. If utility is only ordinal, interpersonal comparisons are meaningless. Nonetheless, economics often adopts utilitarian welfare functions, assuming that comparable utility functions can be calibrated using information beyond consumer choice data. We show that consumer choice data alone are sufficient. As suggested by Edgeworth (1881), just noticeable differences provide a common unit of measure for interpersonal comparisons of utility differences. We prove that a simple monotonicity axiom implies a weighted utilitarian aggregation of preferences, with weights proportional to individual jnd's.

*We thank Paul Milgrom, Philippe Mongin, Uzi Segal, and David Schmeidler for comments and discussions. We are particularly grateful to Luigi Balletta, Ludovic Renou, and an anonymous referee for comments on earlier versions of this paper and for important references. Gilboa gratefully acknowledges support from ISF Grants 204/13, 704/15, the Foerder Foundation, and ERC Grant 269754.

[†]Department of Economics, University of Essex. r_argenziano@essex.ac.uk

[‡]HEC, Paris-Saclay, and Tel-Aviv University. tzachigilboa@gmail.com

1 Introduction

All too often, individuals and societies have to make choices between alternatives that are not Pareto-ranked. Fortunately, in many such problems people seem to share a moral intuition regarding the interpersonal comparison of utilities involved. For example, people tend to view progressive taxation as ethically more desirable than regressive taxation. The reason seems to be that most people would agree that "it is easier" for a rich person to give up, say, $100 than it would be for a poor person (other things being equal). Thus, perhaps judging by introspection, most of us feel that one can compare utility differences across individuals who differ in their wealth or income.

Moreover, people also appear to have shared moral intuition about interpersonal comparisons of utilities across individuals who differ in their perceived needs. Asked whether a disabled person should be allocated a reserved parking spot, most would tend to respond in the affirmative. Similarly, many people would agree that on a crowded bus, a healthy young person should offer their seat to an elderly passenger or a pregnant woman, and not vice versa. When boarding airplanes, airlines give priority to families with young children, allowing them to avoid long lines. The underlying reasoning seems to be that the scarce resource being allocated "matters more" to some individuals than to others. Using remote parking is a smaller burden for a non-disabled person. Standing on a bus is easier for a younger person. Waiting in line is less exhausting for adult passengers.

However, this intuitive notion of utility as a measure of well-being that can be compared across individuals is at odds with the foundation of microeconomic theory. Microeconomics textbooks typically warn the student that utility functions are but mathematical artefacts used to represent preferences. They are shown to be only ordinal and it is emphasised that no particular meaning should be attached to their values or to differences thereof. This realisation, going back to ordinalism of the marginalist revolution (Jevons, 1866, Menger, 1871, Walras, 1874), has been taken to mean that any claim relying on particular properties of a functional form of the utility is meaningless. In particular, in a typical microeconomics course it is considered meaningless to ask which of two individuals would value a good more, whether one's sacrifice is worth the other's benefit, and so forth.

The above notwithstanding, the fields of public economics and social choice have been using social welfare functions, taking individual utility values as arguments, and assuming that these values are meaningful beyond the mere ordering of alternatives for each individual separately. In particular, it is common in economic models to adopt a utilitarian social welfare function (Bentham, 1780), and use a (weighted or unweighted) sum of individuals utilities to formalise a social planner's objective. Examples range from mechanism design (see Borgers, Krahmer and Strausz 2015) to optimal taxation (see Saez and Stantcheva, 2016) and climate change discussions (see Stanton, 2011).

How does social choice theory bridge the gap between the merely-ordinal utility function of consumer theory and the cardinal one needed for much of public economics? One approach follows Harsanyi (1955) in assuming an "impartial observer" who has preferences that are separate from the individuals'. Such an impartial observer would feel that it is easier for the rich to give up income than for the poor, that the frail should get priority seating in public transportation, and so forth. Another approach employs preferences under risk to calibrate individuals' cardinal utilities, and use them in a social choice function (see Harsanyi, 1953, Dhillon and Mertens, 1999, Segal, 2000, Fleurbaey and Mongin, 2016).

While these approaches have their merits, they implicitly agree with the claim that consumer choice data alone do not provide scientific, empirical grounds for interpersonal comparisons of utility.[1] If one believes that economists should only rely on revealed preferences and choice data, the above seems to imply that they should not engage in any normative statements that require interpersonal comparisons of utility.

Our paper points out that this perception is false, because actual consumer choice data, even under certainty, contain much more information than the idealised classroom model assumes. In particular, they contain information that makes it possible, at least in principle, to compare utility differences across individuals on

[1]Sen (1977) and Roberts (1980) classify the different levels of measurability and comparability of utility levels by the degrees of uniqueness of the individual utility functions that each of them requires. Based on their classification, there seems to be a general perception that the assumptions on observables needed for to make interpersonal comparisons and the use of social welfare functions meaningful are rather demanding and largely divorced from the revealed preference paradigm. For example, Myles (2008) writes (p. 54), "Among these alternative degrees of comparability, only ONC [Ordinality and Non-Comparability] and CNC [Cardinality and Non-Comparability] are formally justified by representation theorems on preferences. Moving further down the list [including CUC, Cardinal Unit Comparability, i.e. comparability of gains and losses] requires an increasing degree of pure faith that the procedure is justified."

a scientific basis. Moreover, these comparisons are in line with the common moral intuitions illustrated in the examples we started out with.

Psychological research shows that choice data exhibit limited discernibility[2]: individuals cannot perceive very small differences, and have a positive *just-noticeable-difference* (*jnd*). It follows that indifference is typically an intransitive relation and that consumer choice is more realistically modelled by semi-orders, allowing to capture the fact that close quantities cannot be discerned, than by weak orders, which assume perfect discernibility between any two distinct levels of utility[3].[4] A consumer choice model with semi-ordered – rather than weak-ordered – preferences is not only more realistic, but it also allows for the comparison of utility differences across individuals.

As early as in 1881, Edgeworth suggested to operationalise Bentham's utilitarianism using jnd's. This idea was supported by an axiomatic derivation in Ng (1975). The purpose of the present paper is to contribute to this debate by a different axiomatic derivation of Edgeworth's proposal. Our model and result differ from Ng's in many ways, and we devote the bulk of Section 5 to a discussion of the two. While the results are mathematically independent, we find our axioms significantly easier to accept, and hope that so will some readers.

The basis for Edgeworth's proposal is the observation that cross-individual comparisons of utility differences based on the notion of a jnd capture the common moral intuition in many problems of interest. The intuitive notion that a disabled person needs a parking spot "more" than a healthy one can be captured by jnd's: the former is likely to notice every 100 meters to be crossed, and every step to be climbed, while these might be unnoticeable by the latter. Similarly, in the example of taxation, the vague sense that "it is easier" for the rich to part with $100 than it is for the poor corresponds to the assumption that, at higher levels of income, the

[2] There are many possible sources of data that would be considered observable by psychologists, and that are ignored by the idealized model of consumer theory, including response times, neurological imaging, etc. However, the main point of this paper is that one need not add new *types* of data to make the consumer model more informative than it appears to be in textbooks.

[3] The notion of a just-noticeable-differences (jnd) was introduced and studied by Weber (1834). Luce (1956) suggested to use it as a guideline for the development of consumer theory.

[4] Observe that this means that the standard microeconomics textbook does not only ignore types of data that might exist in reality; it also makes incorrect assumptions about the data it does recognize.

same sum of \$100 "buys less jnd's" than it does at lower levels.[5]

Importantly, we point out that the jnd calculus does not require (or imply) that one would consider absolute levels of utility comparable across individuals. Luckily, making (weighted) utilitarianism operational does not resort to such comparisons either. There is no need to assume that "misery" or "bliss" mean the same thing to different people. Only the *differences* in utilities need to be comparable, and for that jnd's suffice.

Our main contribution is a formal result showing that aggregation of preferences through weighted utilitarianism follows from a very simple axiom. The axiom, which in the context of transitive indifference means nothing but monotonicity, will be shown to simultaneously imply the additively separable structure of utilitarianism, and to select individual weights that reflect people's sensitivities.

While jnd calculus appears intuitive in some examples, it can appear counter-intuitive in others. In particular, some examples call for a richer model, in which time and uncertainty are formally modelled. We briefly describe such an example in Section 4.

This paper is organised as follows. Section 2 illustrates how semi-orders provide a foundation for comparable cardinal utility functions. Section 3 presents the main result of this paper, an axiomatic characterisation of weighted utilitarianism. Section 4 describes an example of an extension of the model. Section 5 discusses the related literature. Section 6 concludes.

2 Intransitive indifference and interpersonal comparisons

2.1 Semi-orders and cardinality

The textbook microeconomic model assumes that preferences are a weak order: a complete and transitive binary relation over alternatives. It follows that a utility function representing such a relation is not unique. It can be replaced by any strictly increasing monotone transformation thereof, without changing the implied preferences. Hence, utility has only ordinal meaning and interpersonal comparisons

[5] In Appendix B we briefly explain Weber's law, which suggests that the jnd is constant on a logarithmic scale.

of utility are meaningless. This conclusion relies on the unrealistic assumption that indifference is transitive.

In reality, choices systematically deviate from transitivity of indifference, due to the limited accuracy of human perception[6]. In the 19th century, the field of psychophysiology observed that a person who is exposed to a physical stimulus would not always notice small changes in it. For example, given two similar masses, a person may not be able to tell which one is larger. Fixing a certain threshold for the probability of identifying the larger quantity (typically 75% in psychological experiments), the *just-noticeable-difference (jnd)* is the minimal increase in the stimulus size that is discernible with probability at the threshold or higher.

Luce (1956) used this observation to refine the model of consumer choice. In a famous example, he argued that one cannot claim to have strict preferences between a cup of coffee without sugar and the same cup with a single grain of sugar added to it. Due to the inability to discern the two, an individual would have to be considered indifferent between them. Similarly, the same individual would most likely be hard pressed to tell which of two cups contains one grain of sugar and which contains two. Indeed, it stands to reason that for small enough grains, an individual would be indifferent between a cup with n grains and one with $(n+1)$ grains of sugar for every n. If transitivity of preferences were to hold, then, by transitivity of indifference, the individual would be indifferent to the amount of sugar in her coffee cup, a conclusion that is obviously false for most individuals. Clearly, the same can be said of any set of alternatives that contain sufficiently close quantities.

Luce (1956) axiomatically defined binary relations, dubbed *semi-orders*, that allow for some types of intransitive indifferences. Appendix B contains more details, including the axioms that Luce proposed to define a semi-order. For the sake of the present discussion, it suffices to say that a semi-order $\succ$ is an irreflexive binary relation $\succ$ (interpreted as strict preference) that can be represented by a utility function u and a threshold $\delta > 0$ (the *utility jnd*) such that, for any two alternatives x, y:

$$x \succ y \qquad iff \qquad u(x) - u(y) > \delta.$$

In words, strict preference emerges for one alternative over another only if the utility difference between them is above the threshold.

[6] See also Mas-Colell, Whinston and Green (1995), p. 7.

A fundamental feature of Luce's model is that utility functions that represent semi-orders are not ordinal; they are "almost unique", hence carry a cardinal meaning. They are almost unique in the following sense: if another function, v, represents $\succ$ as above (with the same $\delta > 0$), u and v have to agree on δ differences. That is, whenever u increases by δ, so should v.[7] More generally, whenever u increases by $k\delta$ for any positive integer k, so should v. For any two alternatives x, y, the functions u, v have to agree not only on the *ordinal ranking* but also on the number of "δ steps" between them, which therefore provides *a cardinal measure of the intensity of preferences.* Consider, for example, three alternatives $x \succ y \succ z$ such that the following holds:

$$\begin{aligned} 4\delta &< u(x) - u(y) \leq 5\delta \\ \delta &< u(y) - u(z) \leq 2\delta. \end{aligned}$$

For any function v representing the same preferences $\succ$ with the same δ, the following must also hold:

$$\begin{aligned} 4\delta &< v(x) - v(y) \leq 5\delta \\ \delta &< v(y) - v(z) \leq 2\delta. \end{aligned}$$

Therefore, for a given consumer it is meaningful to say that "x is better than y *by more than* y is better than z" because the utility jnd provides a scale that is not affected by the choice of utility function[8].

Observe that jnd's are defined by a probability threshold, which is typically taken to be 75%, but can be varied. Thus, one can have a different binary order $\succ_p$ for each probability $p \in [0, 1]$ and assume that each is a semi-order, obtaining a nested family of semi-orders (see Roberts, 1971). Alternatively, one can retain the probabilistic information in its entirety, rather than distinguishing between "higher than p" versus "lower than p", resulting in models of stochastic choice.[9] The main point, however,

[7] Under some richness conditions, one can show that, if both u and v represent the semiorder $\succ$ as above, then there is an increasing function f such that $v = f(u)$, and that, for any a, $f(a + \delta) = f(a) + \delta$.

[8] The utility representation is "almost unique", rather than unique, because for any v and u representing $\succ$, with $v = f(u)$, the function f can be any arbitrary strictly increasing function over the $[0, \delta]$ interval, as long as $f(\delta) - f(0) = \delta$.

[9] See Luce (1958), and, more recently, Gul and Pesendorfer (2006) and many others.

is that even if one restricts attention to commonly-defined jnd's, actual choice data already contain enough information to pin down an almost-unique utility function.

2.2 Interpersonal comparisons of utility

A key step in our analysis is the observation that Luce's model makes interpersonal comparisons of utility meaningful. Denote by $x, y, ...$ social alternatives such as consumption allocations, and assume that each individual i has semi-ordered preferences $\succ_i$ over them, represented by

$$x \succ_i y \qquad iff \qquad u_i(x) - u_i(y) > \delta_i > 0.$$

We have seen in section 2.1 that, given δ_i, the representation of a preference order $\succ_i$ by a function u_i is almost unique. We further notice that a representation by (u_i, δ_i) can be replaced by $(au_i, a\delta_i)$ for any $a > 0$. Without loss of generality, we may select $\delta_i = 1$, that is, replace (u_i, δ_i) by $(\frac{u_i}{\delta_i}, 1)$. The rescaled utilities $\frac{u_i}{\delta_i}$ are directly comparable: For any two consumers $i = 1, 2$, the same increase in $\frac{u_1}{\delta_1}$ and $\frac{u_2}{\delta_2}$ is needed to guarantee that both consumers enjoy one extra jnd. Utility jnd's provide a *common scale* for cardinal utility comparisons.

3 Consistency and weighted utilitarianism

The fact that semi-ordered preferences make interpersonal comparisons of utility meaningful provides a justification for the use of Bergson–Samuelson social welfare functions by a social planner. The next step of our analysis is to provide an axiomatic foundation for a specific social welfare function, that we believe is being implicitly used in the examples discussed in the introduction. Assuming that individual preferences are semi-orders, we investigate what kind of social preferences are consistent with the use of a weighted utilitarian function in which the weight of an individual is the inverse of his jnd. We find that this social welfare criterion is equivalent to assuming that social preferences satisfy a simple form of monotonicity with respect to individual preferences, that we label *Consistency*. We will first present our result, then discuss the implications of the axiom.

The result reported here was inspired by the proof of the main result in Rubinstein (1988), and it is similar to a result in Gilboa and Lapson (1990).[10]

Consider an economy with a set of individuals $N = \{1, \ldots, n\}$. There are $l \geq 1$ goods. Some mathematical details can be simplified if we restrict attention to strictly positive quantities, that is to bundles in $\mathbb{R}^l_{++}$. Assume that individual i's preferences are a semi-order $\succ_i$ on $\mathbb{R}^l_{++}$. Let $\smile_i$ be the reflexive and symmetric relation defined by the absence of $\succ$ in either direction (that is, $x \smile_i y$ if neither $x \succ_i y$ nor $y \succ_i x$). Assume that $\succ_i$ is represented by (u_i, δ_i) as follows: for every $x_i, y_i \in \mathbb{R}^l_{++}$,

$$\begin{aligned} x_i \succ_i y_i \qquad & iff \qquad u_i(x_i) - u_i(y_i) > \delta_i \\ x_i \smile_i y_i \qquad & iff \qquad |u_i(x_i) - u_i(y_i)| \leq \delta_i \end{aligned} \tag{1}$$

We assume that u_i is weakly monotone and continuous, and that $\delta_i > 0$. We will also assume that, for each i, $\succ_i$ is *unbounded*: for every $x_i \in \mathbb{R}^l_{++}$, there exist $y_i, z_i \in \mathbb{R}^l_{++}$ such that $y_i \succ_i x_i \succ_i z_i$. The representation (1) implies that u_i is unbounded, and its continuity implies that $image\,(u_i) = \mathbb{R}$.

An allocation is an assignment of bundles to individuals,

$$x = (x_1, ..., x_n) \in X \equiv \left(\mathbb{R}^l_{++}\right)^n.$$

Assume that society has semi-ordered[11] preferences $\succ_0$ on the set of allocations $\left(\mathbb{R}^l_{++}\right)^n$ that is represented by (u_0, δ_0) with $\delta_0 > 0$. Without loss of generality, let $\delta_0 = 1$. Thus, $u_0 : \left(\mathbb{R}^l_{++}\right)^n \to \mathbb{R}$ is such that, for every $x, y \in \left(\mathbb{R}^l_{++}\right)^n$,

$$\begin{aligned} x \succ_0 y \qquad & iff \qquad u_0(x) - u_0(y) > 1 \\ x \smile_0 y \qquad & iff \qquad |u_0(x) - u_0(y)| \leq 1 \end{aligned} \tag{2}$$

We similarly assume that u_0 is continuous.

For $z \in X$ and $x_i \in \mathbb{R}^l_+$ we denote by $(z_{-i}, x_i) \in X$ the allocation obtained by

[10] Rubinstein (1988) dealt with procedures for choices under risk. While his monotonicity condition cannot apply in the current set-up, his proof relies on an insight that proved useful also in Gilboa and Lapson (1990). The latter contained two interpretations of a main result, one for decision under uncertainty and one for social choice. In the published version (1995) only the former appeared. The result presented here differs from those of Gilboa and Lapson (1990, 1995) in a number of mathematical details.

[11] We will discuss the implications of assuming that social preferences are given by a standard weak order at the end of this section.

replacing the i-th component of z, z_i, by x_i. The axiom we will use to characterise weighted utilitarianism is

Consistency: For every i, every $z \in X$ and every $x_i, y_i \in \mathbb{R}^l_{++}$,

$$(z_{-i}, x_i) \succ_0 (z_{-i}, y_i) \qquad iff \qquad x_i \succ_i y_i.$$

First, observe that if all jnd's were zero, Consistency would boil down to simple monotonicity of society's preferences with respect to the individuals': if all individuals' bundles apart from i stay fixed, society adopts i's preferences.[12]

In the presence of semi-ordered preferences, Consistency still states that if we focus on an individual i, and hold all other individuals' bundles fixed, society's preferences are those of the individual. In case individual i expresses strict preference, say $x_i \succ_i y_i$, society agrees with that individual. Similarly, if individual i cannot tell the difference between x_i and y_i, the difference between the two is immaterial to society as well.

Notice that Consistency is restricted to the case that no individual $j \neq i$ is affected at all, whether she can tell the difference or not, i.e., that z_{-i} is kept exactly constant when comparing (z_{-i}, y_i) to (z_{-i}, x_i). Importantly, Consistency does *not* require that society agrees with i's preferences as long as this individual is the only one to express strict preference, while the others might be affected by the choice in a way they cannot discern. For example, consider a suggestion that each individual $j \neq i$ contribute 1 cent to i. Assume that 1 cent is a small enough quantity for each $j \neq i$ not to notice it. By contrast, the accumulation of these cents can render i rich. Consistency does *not* imply that society prefers this donation scheme. Indeed, requiring this implication would lead to intransitivities (as one can change the happy recipient of the individually-negligible donations and generate cycles of strict societal preferences). Beyond intransitivity, such a version of the axiom does not seem well fit to capture social preferences with which people would agree. In the example above, the two bundles differ for all consumers. Individual i can tell the difference between the alternatives, while the individuals who lose one cent cannot. Yet, such a donation scheme has a feeling of deception, involving as it were fine print which is unnoticeable at a given time but may prove noticeable in the long

[12] See also Fleming (1952), who derives a utilitarian aggregation in the context of standard preferences, assuming the trade-off between any pair of individuals is independent of the others.

run. Thus, such a donation scheme might not capture common moral sentiments. Importantly, Consistency does not resort to such schemes. It only applies when the bundles of all the individuals, but one, are unaffected.

For the statement of the main result we need the following definition: a *jnd-grid* of allocations is a collection $A \subset X$ such that, for every $x, y \in A$ and every $i \in N$,

$$u_i(x_i) - u_i(y_i) = k_i \delta_i \qquad \text{for some } k_i \in \mathbb{Z}$$

Thus, a jnd-grid is a countable subset of allocations, such that the utility differences between any two elements thereof, for any individual, is an integer multiple of that individual's jnd.

We can now state

Theorem 1 *Let there be given $(\succ_i)_{i \in N}, ((u_i, \delta_i))_{i \in N}, \succ_0$ and u_0 as above. Consistency holds iff there exists a strictly monotone, continuous*

$$g : \mathbb{R}^n \to \mathbb{R}$$

such that for every $x \in X$

$$u_0(x) = g(u_1(x_1), ..., u_n(x_n))$$

and, for every jnd-grid $A \subset X$ there exists $c \in \mathbb{R}$ such that, for every $x \in A$,

$$u_0(x) = c + \sum_{i=1}^{n} \frac{1}{\delta_i} u_i(x_i) \tag{3}$$

The theorem states that, should society's preferences satisfy Consistency with respect to the individuals' preferences, the former should basically be represented by a weighted (utilitarian) summation of the individuals' utilities, where the weights are the inverse of the just noticeable differences. This is the version of utilitarianism first suggested by Edgeworth (1881, p. 60) , who wrote, "Just perceivable increments of pleasure, of all pleasures for all persons, are equateable." In fact, given individual preferences $\succ_i$ represented by (u_i, δ_i), the social welfare function (3) considers the equivalent representation $\left(\frac{u_i}{\delta_i}, 1\right)$, in which jnd's have been equated, and gives equal weights to all individuals.

We now discuss in more detail the implications of the Consistency axiom. The "if" part of the axiom is a rather natural form of monotonicity. Consider two allocations (z_{-i}, x_i) and (z_{-i}, y_i). In case individual i expresses strict preference, say $x_i \succ_i y_i$, there seems to be no reason for society not to agree with that individual, as no one else is affected by the choice.

The "only if" part of the axiom is more interesting. It states that society is no more sensitive than the individual herself. If individual i cannot tell the difference between x_i and y_i, the difference between the two is immaterial to society as well. A useful way to think about this is to interpret $x_i \smile_i y_i$ and $x \smile_0 y$ as "too close to be worth worrying about".[13] If it is the case that only one individual is allotted a different bundle under x as compared to y, and this individual doesn't find the difference of significance, it seems reasonable that neither would society. Notice that this rules out the possibility that social preferences are given by a standard weak order instead of a semi-order.

The implications of the "only if" part of the axiom on the characterisation is twofold. First, it implies that individuals should be compared by their jnd's. Second, and probably more surprisingly, it implies the additive structure of the welfare function. One need not assume a utilitarian function as did Bentham – one obtains it from the axiom. Without the "only if" part of the axiom, many more social welfare functions would represent social preferences $\succ_0$. For example, one could consider a vector of positive weights $\lambda = (\lambda_i)_i$ and define the social welfare function

$$u_0^\lambda(x) = c + \sum_{i=1}^{n} \frac{\lambda_i}{\delta_i} u_i(x_i).$$

For any weights $\lambda_i \geq 1$, the preferences defined by this function would satisfy the "if" part of Consistency, but u_0^λ would allow a variety of weights of the individuals. Most importantly, functional forms that are not additively separable would also be compatible with this weaker version of the axiom.

One may replace the "only if" part of the axiom by assuming that social preferences are given by a weak order, and add an "anonymity" condition, stating that if one individual's bundle can be replaced by another bundle that is precisely one jnd better for that individual, society is indifferent regarding the identity of the

[13] Rounding off tax returns to the closest integer amount may be considered an example in which some differences are considered to be below the jnd for a particular individual and for society.

individual. Indeed, the fact that for weak-ordered social preferences Consistency is stated as an *if and only if* condition implicitly requires that society treat the jnd's of different individuals equally. However, the axiom does not explicitly refer to more than one jnd, and or to separability across individuals. It is a result of the theorem that, with semi-ordered preferences for the individuals, this axiom suffices to derive additive separability across individuals.

We claim that the welfare function axiomatised in Theorem 1, where an individual's weight is the inverse of their jnd, captures the common moral feeling underlying the examples in the introduction. Giving a higher weight to a person with a smaller jnd means being more likely to allocate a good to someone for whom it creates more jnd's. That it, to the person to whom it "matters more", which is what most people would agree should be done when, for example, choosing whether to allocate a parking spot to a disabled or a non-disabled driver, a priority seating for the aged, and so forth.

Similarly, the Edgeworth version of weighted utilitarianism seems to capture the social preferences underlying commonly shared ethical opinions about taxation. People tend to view progressive taxation as ethically more desirable than, say, regressive taxation, because they feel that it is "easier" for a rich person to pay a given amount, than it would be for a poor person (other things being equal). This is normally explained by arguing that the marginal utility of money is decreasing. In Luce's model, one can provide empirical meaning to such a claim. Suppose that the alternatives are real-valued, denoting the cost (say, in dollars) of a bundle one may consume a day. The value 0 denotes destitution, implying starvation. The value 1 allows one to consume a loaf of bread, clearly a very noticeable difference. In fact, even the value 0.1, denoting the amount of bread one can buy for 10 cents, is noticeably different from 0 for a starving person. However, when one's daily consumption is a bundle that costs \$500, it is unlikely that a bundle that costs \$501 would make a large enough difference to be noticed. Thus, when starting at 0, the first dollar makes a noticeable difference, but the 500th does not. More generally, there probably are more jnd's between the bundle bought at \$100 and the empty bundle than there are between the bundle bought at \$200 and the former; that is, the "second \$100 buys one less jnd's than the first \$100".

4 Time and Endogeneity

In contrast to the examples discussed above, there are others in which using people's sensitivities as a measure of their importance, so to speak, strikes most of us as counter-intuitive if not blatantly outrageous. For example, assume that we need to divide two bottles of wine between two individuals. The wines differ in their quality, one being exquisite according to wine experts, and the other not. It so happens that the individuals also differ: one of them is a wine connoisseur and the other isn't. The connoisseur sees many jnd's between the wines, while the layperson doesn't. Thus, Edgeworth solution would be to give the better wine to the expert and let the layperson make do with the lesser wine.

The reason that the Edgeworth solution appears unfair in the wine example is that, in the back of our minds, we believe that the layperson can also be educated and develop a more refined taste. Worse still, such education would typically require consumption. It thus appears as if Edgeworth's suggestion allows the connoisseur, who might have developed refined tastes through consumption thanks to being rich, will be allowed to consume more of the good wine than the poor, who has not had a chance to consume and learn to appreciate quality.

The possibility that taste might change, becoming more or less refined, calls for a more elaborate model. Consider the wine example, with two goods and two periods. There are two individuals, $i = 1, 2$, each of whom consumes, at time $t = 1, 2$, quantity x_i^t of the lower quantity wine and quantity w_i^t of the higher quantity wine. Assume that individual i's preferences are represented by $(u_i, 1)$, with

$$\begin{aligned} u_i\left(x_i^1, w_i^1, x_i^2, w_i^2\right) &= \log\left(1 + x_i^1\right) + A_i \log\left(1 + w_i^1\right) \\ &\quad + \log\left(1 + x_i^2\right) + \left[A_i + B_i\left(1 - e^{-w_i^1}\right)\right] \log\left(1 + w_i^2\right) \end{aligned}$$

with the following interpretation: both individuals derive the same utility from the lower quality wine, $\log(1 + x)$, in both periods.[14] As for the higher quality one, they each begin with a level of sensitivity $A_i \geq 0$, so that the quantity of high quality wine each of them consumes generates more jnd's the higher is A_i.

Note that if there were only one period $t = 1$, the individual with the higher

[14] Note that we normalized each (u_i, δ_i) so that the weights of the individuals in the SWF are identical; that is, the Bentham-Edgeworth suggestion is to maximize $u_1 + u_2$, and this would be the conclusion of our theorem on jnd-grids as well.

sensitivity would indeed be allocated a higher quantity of the high quality wine. However, the utility at time $t = 2$ depends also on the consumption at time $t = 1$: the individuals are assumed to refine their taste as a result of consumption, where $B_i \geq 0$ is interpreted as individual i's ability to acquire taste. With no past consumption, $w_i^1 = 0$, individual i is left with her initial sensitivity A_i, but as her past consumption grows her sensitivity increases as well (converging to $A_i + B_i$ as $w_i^1 \to \infty$). In this case, maximisation of the utilitarian social welfare function would require that the less sensitive individual consume some of the high quality wine already in the first period, so that in the second period she can "contribute" more jnd's to the SWF. That is, it is worth investing in refining individual i's taste, even at the cost of reducing the current jnd's experienced by individual $j \neq i$, so that in the next period society would have two "jnd-producing engines" rather than one. Clearly, the argument will be stronger the more period await the individuals in the future, that is, if we allow $T > 2$ periods.[15]

Along similar lines, one may add to the model several states of the world, allowing for the possibility that there is uncertainty about acquiring taste. For example, it is possible that only a certain percentage of children would acquire a taste for certain types of consumption (say, forms of art). Still, the jnd's that these children will accumulate throughout their lifetime might justify the investment of resources (a net loss of jnd's) at present. Thus, the model presented above is highly idealised. In many examples one would need to extend it to multiple periods and multiple states of the world. When these are taken into account, it doesn't strike us as counter-intuitive to suggest that society "collect" jnd's across individuals, periods, and states.

5 Related Literature

5.1 Other data sources

The bulk of the literature in public economics and in social choice theory, when in need of a social welfare function, tends to assume that such a function is simply given, or that it is derived from observables that go beyond individual choice.

[15] One may also extend the model to the past, arguing the individual differences in the coefficients A_i are themselves a result of past consumption rather than innate.

For example, Harsanyi's impartial observer (Harsanyi, 1955) is assumed to have preferences that are separate from the individuals'. We have no quarrel with this assumption on philosophical grounds. However, methodologically it appears to introduce a gap between the standard microeconomic foundations of consumer theory and social choice. Our main point is that this apparent gap can be bridged simply be recognising that actual preferences are more informative than normally assumed. Indeed, the impartial observer's preferences may well be consistent with our jnd calculus, and we hold that in many examples they are.

Another approach suggests pinning down a cardinal utility function for each individual based on their preferences under risk, and then using these functions, with some normalisation, for the social welfare function (see Harsanyi (1953), Dhillon and Mertens (1999), Fleurbaey and Mongin (2016), Fleurbaey and Zuber (2017)). We do not discuss here the philosophical underpinnings of this approach.[16] Segal (2000) proposes (and axiomatically derives) a weighted utilitarian solution in which, when evaluating an alternative, each individual is assigned a weight which is inversely proportional to that individual's gain in von Neumann-Morgenstern utility that this alternative promises, relative to a benchmark. While this solution differs from the one discussed here in several way, both mathematical and conceptual, the two share a fundamental intuition, according to which the less fortunate should have a higher weight in the social welfare function.

The main point of the present paper is that one need not add information to the consumer model (such as preferences over lotteries) in order to render utilitarianism meaningful. In fact, all that one needs to do is to be somewhat more realistic about the data that consumer choices actually offer. It is possible to go much further than jnd's, and to introduce into the model probabilities of choice (beyond the 75% probability threshold); response time; self-report; brain activity; etc. While these additional sources of data can be introduced, they are not necessary in order to bridge the gap between microeconomics textbook and economic practice.

5.2 Ng (1975)

We now turn to explain what we view as the marginal contribution of our paper relative to Ng (1975), which provides a different axiomatic derivation of a utilitarian

[16] These include, inter alia, the introduction of risk attitudes into the ethical discussion and the additional assumptions about the range of the utility values of each individual.

welfare function with utility weights given by the inverse of jnd's. The key difference between the two derivations is in the main assumption: the "Majority Weak Preference (MWP) Criterion" in Ng (1975) and "Consistency" above. MWP states that, if there is a weak majority of individuals who strictly prefer one alternative over another, while no individual has the opposite preference, society should respect the majority preference. Several related points emerge in comparing the axioms:

a. The MWP axiom involves counting individuals. The theorem shows that adding up individuals in each of two sets (those who prefer an alternative x to y and those who don't) eventually leads to adding up the utilities of the individuals. This is far from trivial, but from a conceptual viewpoint it feels as if addition is explicitly assumed.

b. The MWP axiom involves counting by size of the utility difference. To identify the two sets of individuals to be counted (those who prefer an alternative x to y and those who don't), the axiom requires to count how many utility differences are above the individual jnd and how many are below it[17] – and thus the intuition behind the axiom seems to directly appeal to the additive form we would like to derive.

c. The MWP axiom allows society to penalise an individual in a way that favours others, as long as this is unnoticeable by the individual. Suppose there are only two individuals, and consider a donation scheme similar to the one we discussed in section 3: one cent is transferred from j to i. Assume this makes a noticeable difference for i, but not for j. The MWP axiom implies that society should approve such a transfer, because one jnd is being gained and none is being lost. As we discussed in section 3, such a transfer involves a feeling of deception, and assuming that society approves it might not capture common moral sentiments.

In comparison to MWP, the Consistency axiom (i) makes no summation over individuals (ii) makes no implicit additions of utility differences; (iii) does not involve reducing people's underlying utility functions in unnoticeable ways. The assumption that society's jnd on each axis is equal to that of the individual in question does indeed make an implicit comparison of jnd's, saying that increasing one individual's underlying utility by one jnd is equivalent to doing this with another's. But no counting of individuals is involved; no comparisons of sets of differences (that is, counting how many are below and how many are above their jnd); and no "cheating" in terms of "shaving off" unnoticeable utility differences from some individuals.

[17] Indeed, without this restriction the axiom would lead to cycles of strict preferences.

Several further differences between our model and Ng's (1975) exist, but they are rather minor, and may not be worth dwelling upon. These include:

– Ng assumes that the Social Welfare Function (SWF) W is quasi-convex or quasi-concave as a function of the individual utilities, U_i, while we make no such assumption. The condition is needed in Ng's theorem only to solve the problem that the SWF can be nonlinear "in the small", namely to behave in weird ways within "boxes" defined by one jnd per individual. We solve this problem by discussing only jnd-grids.

– Ng assumes that the SWF W is a function of the individual utility functions, U_i's, and that this function is continuous and differentiable. We deduce that W is a continuous function of the U_i's as a result of the Consistency axiom, and continuity of W with respect to the underlying allocations – as one would expect of a monotonicity condition – and we don't need differentiability.

– The current paper assumes that society has a semi-order, rather than a standard weak order. As mentioned above, this can be modified, but at the cost of an additional assumption, saying that, starting with two equivalent allocations, increasing one of two individuals' utility by their jnd results in two equivalent allocations as well.

6 Conclusions

The main conceptual message of our paper is that one need not have additional sources of data beyond choice data to make the interpersonal comparison of utility differences meaningful. In fact, scholars who wish to adhere to the revealed preference paradigm as strictly as possible will find that choice data provide evidence of intransitive indifference.

Despite Ng's (1975) contribution, this point seems to be generally ignored. As cited above, Myles writes (2008, p. 54), "Among these alternative degrees of comparability, only ONC [Ordinality and Non-Comparability] and CNC [Cardinality and Non-Comparability] are formally justified by representation theorems on preferences. Moving further down the list [including CUC, Cardinal Unit Comparability, i.e. comparability of gains and losses] requires an increasing degree of pure faith that the procedure is justified."

One can only speculate about the reasons that this point hasn't been accepted by the profession following Edgeworth (1881) and Ng (1975). Edgeworth's proposal was made, perhaps, too early, as it predated the adoption of the revealed preference paradigm by several decades. Ng (1975), by contrast, relies on semi-orders and cites their observable foundations. However, Ng (1975, and in other publications) was willing to make interpersonal comparisons of absolute utility levels as well (see his discussion on p. 558, second paragraph, beginning with "The only logical…"). In particular, Ng is willing to compare "absolute bliss" or "absolute misery" across individuals. Many economists might be wary of such comparisons, as individuals' personal experiences or religious beliefs might differently colour the meaning of such concepts to them.

It is possible that these interpersonal comparisons of absolute levels of utility beclouded the point that we find essential: the claim that Cardinal Unit Comparisons are divorced from empirical data is only true within the highly idealised model of economics textbook. In reality no additional information beyond choice data are needed to make such comparisons.

Appendix A: Proof of Theorem 1

First, assume Consistency. Some of the arguments until and including Claim 7 are rather standard, and similar arguments appear in the literature (see, for instance, Blackorby, Primont, and Russell, 1978, and Fleurbaey and Mongin, 2016). Yet, we have not found references that are sufficiently close to our set-up to cite without proof, and we provide the proofs for completeness.

Claim 1: For every $z \in X$ and every $i \leq n$, we have $image\left(u_0\left(z_{-i}, \cdot\right)\right) = \mathbb{R}$.

Proof: Fixing i and z_{-i}, Consistency implies that the social preference $\succ_0$ is dictated by $\succ_i$. Hence it is unbounded: for every $x_i \in \mathbb{R}^l_{++}$ there are $y_i, w_i \in \mathbb{R}^l_{++}$ such that $(z_{-i}, y_i) \succ_0 (z_{-i}, x_i) \succ_0 (z_{-i}, w_i)$. This implies that $image\left(u_0\left(z_{-i}, \cdot\right)\right)$ is unbounded (from below and from above). Given that u_0 is continuous, its range is also convex, and $image\left(u_0\left(z_{-i}, \cdot\right)\right) = \mathbb{R}$ follows. □

Claim 2: For every $z \in X$, every $i \leq n$, and every $x_i, y_i \in \mathbb{R}^l_{++}$, if $u_i\left(x_i\right) \geq u_i\left(y_i\right)$, then $u_0\left(z_{-i}, x_i\right) \geq u_0\left(z_{-i}, y_i\right)$.

Proof: Assume that this is not the case for some z, i, x_i, y_i. Then we have $u_i\left(x_i\right) \geq u_i\left(y_i\right)$ but $u_0\left(z_{-i}, x_i\right) < u_0\left(z_{-i}, y_i\right)$. By Claim 1 we can find $w_i \in \mathbb{R}^l_{++}$

such that

$$u_0(z_{-i}, x_i) - 1 < u_0(z_{-i}, w_i) < u_0(z_{-i}, y_i) - 1$$

so that

$$u_0(z_{-i}, x_i) < u_0(z_{-i}, w_i) + 1 < u_0(z_{-i}, y_i)$$

It follows that $(z_{-i}, y_i) \succ_0 (z_{-i}, w_i)$ but it is not the case that $(z_{-i}, x_i) \succ_0 (z_{-i}, w_i)$. By Consistency, this implies that $y_i \succ_i w_i$ but not $x_i \succ_i w_i$. This, however, is impossible as the first preference implies $u_i(y_i) > u_i(w_i) + \delta_i$, which implies $u_i(x_i) > u_i(w_i) + \delta_i$, which, in turn, could only hold if $x_i \succ_i w_i$ were the case. □

Claim 3: For every $z \in X$, every $i \leq n$, and every $x_i, y_i \in \mathbb{R}^l_{++}$, if $u_i(x_i) > u_i(y_i)$, then $u_0(z_{-i}, x_i) > u_0(z_{-i}, y_i)$.

Proof: Assume that z, i, x_i, y_i are given with $u_i(x_i) > u_i(y_i)$. As $image(u_i) = \mathbb{R}$ we can find $w_i \in \mathbb{R}^l_{++}$ such that

$$u_i(z_{-i}, y_i) < u_i(z_{-i}, w_i) + \delta_i < u_i(z_{-i}, x_i)$$

so that $x_i \succ_i w_i$ but not $y_i \succ_i w_i$. By Consistency, $(z_{-i}, x_i) \succ_0 (z_{-i}, w_i)$ but not $(z_{-i}, y_i) \succ_0 (z_{-i}, w_i)$. The first preference implies $u_0(z_{-i}, x_i) > u_0(z_{-i}, w_i) + 1$ while the second $u_0(z_{-i}, y_i) \leq u_0(z_{-i}, w_i) + 1$. Hence $u_0(z_{-i}, x_i) > u_0(z_{-i}, y_i)$ follows. □

Claim 4: For every $x, y \in X$, if for every $i \leq n$, $u_i(x_i) \geq u_i(y_i)$, then $u_0(x) \geq u_0(y)$.

Proof: Use Claim 2 inductively. □

Claim 5: There exists a function $g : \mathbb{R}^n \to \mathbb{R}$ such that for every $x \in X$

$$u_0(x) = g(u_1(x_1), ..., u_n(x_n)).$$

Proof: We need to show that, for every $x, y \in X$, if for every $i \leq n$, $u_i(x_i) = u_i(y_i)$, then $u_0(x) = u_0(y)$. This follows from using Claim 4 twice. □

Claim 6: The function $g : \mathbb{R}^n \to \mathbb{R}$ is strictly monotone.

Proof: This follows from Claims 4 and 5. □

Claim 7: The function $g : \mathbb{R}^n \to \mathbb{R}$ is continuous.[18]

[18] See also Lemma 1 (p. 296) in Fleurbaey and Mongin (2016).

Proof: Assume it were not. Then there would be a point of discontinuity $\alpha = (\alpha_1, ..., \alpha_n) \in \mathbb{R}^n$. In particular, there would be a sequence $\alpha^k \in \mathbb{R}^n$ for $k \geq 1$ such that $\alpha^k \rightarrow_{k\rightarrow\infty} \alpha$ but $g(\alpha^k)$ does not converge to $g(\alpha)$. That is, there exists $\varepsilon > 0$ such that there are infinitely many k's for which $g(\alpha^k) < g(\alpha) - \varepsilon$ or there are infinitely many k's for which $g(\alpha^k) > g(\alpha) + \varepsilon$. Assume without loss of generality that it is the former case, and that $g(\alpha^k) < g(\alpha) - \varepsilon$ holds for every k.

Because $image(u_i) = \mathbb{R}$ for every i, we can find $x_i \in \mathbb{R}^l_{++}$ such that $u_i(x_i) = \alpha_i$. We wish to construct a sequence $x_i^k \in \mathbb{R}^l_{++}$ for each i such that $u_i(x_i^k) = \alpha_i^k$ and that $x_i^k \rightarrow_{k\rightarrow\infty} x_i$. If such a sequence existed, we would have $x^k = (x_1^k, ..., x_n^k) \rightarrow_{k\rightarrow\infty} x$ while

$$\begin{aligned} u_0(x^k) &= g(u_1(x_1^k), ..., u_n(x_n^k)) = g(\alpha^k) \\ &< g(\alpha) - \varepsilon \\ &= g(u_1(x_1), ..., u_n(x_n)) - \varepsilon = u_0(x) - \varepsilon \end{aligned}$$

for every k, contradicting the continuity of u_0.

Consider, then $i \leq n$ and $k \geq 1$. Let

$$A_i^k = \left\{ w_i \in \mathbb{R}^l_{++} \,\middle|\, u_i(w) = \alpha_i^k \right\}.$$

As $image(u_i) = \mathbb{R}$, $A_i^k \neq \varnothing$. Because u_i is continuous, A_i^k is closed. Hence there exists a closest point $w_i \in A_i^k$ to x_i. (To see this, choose an arbitrary point $w_i \in A_i^k$ and consider the intersection of A_i^k with the closed ball around x_i of radius $\|w_i - x_i\|$.) Choose such a closest point $x_i^k \in A_i^k$ for each i.

We claim that x_i^k converge to x_i. Let there be given $\varsigma > 0$. Consider the ς-ball around x_i, $N_\varsigma(x_i)$. Due to strict monotonicity, u_i obtains some value $\beta_i < \alpha_i$ as well as some other value $\gamma_i > \alpha_i$ on $N_\varsigma(x_i)$, and, by continuity, the range of u_i restricted to $N_\varsigma(x_i)$ contains the entire interval $[\beta_i, \gamma_i]$. As $\alpha_i^k \rightarrow_{k\rightarrow\infty} \alpha_i$, for large enough k's $\alpha_i^k \in [\beta_i, \gamma_i]$ and one need not look beyond $N_\varsigma(x_i)$ to find a point $w_i \in A_i^k$. In other words, for large enough k's, $x_i^k \in N_\varsigma(x_i)$ and $x_i^k \rightarrow_{k\rightarrow\infty} x_i$ follows. This completes the proof of continuity of g. $\square$

To complete this part of the proof we wish to show that for every jnd-grid $A \subset X$

there exists $c \in \mathbb{R}$ such that, for every $x \in A$,

$$u_0(x) = c + \sum_{i=1}^{n} \frac{1}{\delta_i} u_i(x_i).$$

To this end we state

Claim 8: For every $\alpha \in \mathbb{R}^n$ and every $i \leq n$,

$$g(\alpha + \delta_i \mathit{1}_i) = g(\alpha) + 1$$

(where $\mathit{1}_i$ is the i-th unit vector).

Proof: Consider $\alpha \in \mathbb{R}^n$ and $x_i \in \mathbb{R}^l_{++}$ such that $u_i(x_i) = \alpha_i$. Let $y_i \in \mathbb{R}^l_{++}$ be such that $u_i(y_i) = \alpha_i + \delta_i$. Then it is not the case that $y_i \succ_i x_i$ and, by Consistency, it is also not the case that $(x_{-i}, y_i) \succ_0 x$. Hence, $u_0(x_{-i}, y_i) \leq u_0(x) + 1$ and $g(\alpha + \delta_i \mathit{1}_i) \leq g(\alpha) + 1$ follows.

Next, for every $k \geq 1$, we can pick $y_i^k \in \mathbb{R}^l_{++}$ be such that $u_i(y_i) = \alpha_i + \delta_i + \frac{1}{k}$. Then $y_i^k \succ_i x_i$ and, by Consistency again, $\left(x_{-i}, y_i^k\right) \succ_0 x$, implying $u_0\left(x_{-i}, y_i^k\right) > u_0(x) + 1$ and $g\left(\alpha + \left(\delta_i + \frac{1}{k}\right) \mathit{1}_i\right) > g(\alpha) + 1$. By continuity of g, this implies $g(\alpha + \delta_i \mathit{1}_i) \geq g(\alpha) + 1$.

Combining the two, $g(\alpha + \delta_i \mathit{1}_i) = g(\alpha) + 1$ follows. □

Claim 9: For every jnd-grid $A \subset X$ there exists $c \in \mathbb{R}$ such that, for every $x \in A$,

$$u_0(x) = c + \sum_{i=1}^{n} \frac{1}{\delta_i} u_i(x_i).$$

Proof: Pick an arbitrary $x \in A$ to determine the value of c, and proceed by inductive application of Claim 8 (over the countable jnd-grid). □

This completes the sufficiency of Consistency for the existence of the function g with the required properties. We now turn to the converse direction, that is, the necessity of Consistency. Assume, then, that there exists a strictly monotone, continuous $g : \mathbb{R}^n \to \mathbb{R}$ such that for every $x \in X$ $u_0(x) = g(u_1(x_1), ..., u_n(x_n))$ and, for every jnd-grid $A \subset X$ there exists $c \in \mathbb{R}$ such that, for every $x \in A$,

$$u_0(x) = c + \sum_{i=1}^{n} \frac{1}{\delta_i} u_i(x_i).$$

To prove Consistency, let there be given $i \leq n$, $z \in X$ and $x_i, y_i \in \mathbb{R}^l_{++}$. We need to show that $(z_{-i}, x_i) \succ_0 (z_{-i}, y_i)$ holds iff $x_i \succ_i y_i$. Assume first that $(z_{-i}, x_i) \succ_0 (z_{-i}, y_i)$. Then $u_0((z_{-i}, x_i)) > u_0((z_{-i}, y_i)) + 1$. Consider the jnd-grid A that contains (z_{-i}, x_i). Let $w_i \in \mathbb{R}^l_{++}$ be such that $u_i(w_i) = u_i(x_i) - \delta_i$, so that $(z_{-i}, w_i) \in A$. It follows that $u_0((z_{-i}, w_i)) = u_0((z_{-i}, x_i)) - 1$. Note that

$$u_0((z_{-i}, y_i)) < u_0((z_{-i}, x_i)) - 1 = u_0((z_{-i}, w_i)).$$

By monotonicity of g, this can only hold if

$$u_i(y_i) < u_i(w_i) = u_i(x_i) - \delta_i$$

and $x_i \succ_i y_i$ follows.

Conversely, if $x_i \succ_i y_i$ holds, we can find $w_i \in \mathbb{R}^l_{++}$ be such that $u_i(w_i) = u_i(x_i) - \delta_i > u_i(y_i)$ and show that $u_0((z_{-i}, w_i)) = u_0((z_{-i}, x_i)) - 1$ while $u_0((z_{-i}, w_i)) > u_0((z_{-i}, y_i))$ so that $u_0((z_{-i}, x_i)) - 1 > u_0((z_{-i}, y_i))$ and $(z_{-i}, x_i) \succ_0 (z_{-i}, y_i)$ follows. $\square\square\square$

Appendix B: Just-Noticeable Differences and Semi-Orders

Back in the early 19th century the field of psychophysiology studied mechanisms of discernibility that cast a dark shadow of doubt on the neoclassical model. Weber (1834) asked, what is the minimal degree of change in a stimulus needed for this change to be noticed. For example, holding two ores, one weighing S grams and the other – $(S + \Delta S)$ grams, a person will not always be able to tell which is heavier. To be precise, when ΔS is zero the person's guess would be expected to correct 50% of the time. As ΔS goes to infinity, the chance of missing the larger weight goes to zero. Fixing a probability threshold – commonly, at 75% – one may ask what the minimal ΔS that reaches that threshold is, and how it behaves as a function of S. *Weber's law* states that this threshold behaves proportionately to S. That is, there exists a constant $C > 1$ that

$$(S + \Delta S)/S = C.$$

Thus, if the base-level stimulus is multiplied by a factor $a > 0$, the minimal change required to be noticed (with the same threshold probability) is $a\Delta S$. Equivalently, a change ΔS will be noticed only if

$$\log(S + \Delta S) - \log(S) > \delta \equiv \log(C) > 0. \tag{4}$$

This law is considered a rather good first approximation and it appears in most introductory psychology textbooks.[19]

Luce (1956) used this observation to refine the model of consumer choice. Luce defined binary relations that allow for some types of intransitive indifferences. A *semi-order* is an irreflexive binary relation $\succ$ (interpreted as strict preference) that satisfies two axioms. To state them, let $\smile$ be the reflexive and symmetric relation defined by the absence of $\succ$ in either direction (that is, $x \smile y$ if neither $x \succ y$ nor $y \succ x$). The axioms are:

[19] It is often mentioned in the context of the Weber-Fechner law. Fechner (1860) was interested also in subjective perception. Over the past decades, Stevens's power law is considered to be a better approximation of subjective perceptions than is Fechner's law. However, as far as discernibility is concerned, Weber's law probably still holds the claim to be the best first approximation. See Algom (2001).

L1. If $x \succ y$ and $y \succ z$ but $z \smile w$, then $x \succ w$;

L2. If $x \succ y$ and $z \succ w$ but $y \smile z$, then $x \succ w$.

It is easy to verify that Pareto domination (in the role of strict preference $\succ$) satisfies none of these axioms. Consider L1. It is possible that w is preferred by one individual to each of $\{x, y, z\}$ and that the converse is true for another individual. In that case, z and w will be incomparable for a fundamental reason, and so will be x and w. L1 rules that out. In a sense, it suggests that the incomparability of z and w can only be due to their proximity on the utility scale, and, given that $x \succ y \succ z$, this proximity cannot hold for x and w. Similar reasoning applies to L2.

In the case of a finite set of alternatives, Luce proved that $\succ$ is a semi-order if and only if it can be represented by a pair (u, δ) where u is a utility function on the set of alternatives and $\delta > 0$ is a threshold – called the *just noticeable difference* (jnd) – such that, for every x, y,

$$x \succ y \qquad iff \qquad u(x) - u(y) > \delta \tag{5}$$

$$x \smile y \qquad iff \qquad |u(x) - u(y)| \leq \delta \tag{6}$$

If the set of alternatives is infinite, additional conditions are required for the representation above. We will discuss only semi-orders that have such a representation.[20]

Observe that the indifference relation $\smile$ contains two types of pairs: alternatives (x, y) that are too similar to each other to be told apart, as in the case of a single dimension, but also alternatives that are clearly discernible from each other but are consciously considered to be equivalent. The representation (5) (and the implied (6)) suggests that the utility function would map all "conscious" equivalences onto sufficiently close points on the real line, so that both reasons for indifference – indiscernibility and equivalence – are mapped into proximity of the utility values.

Given a semi-order $\succ$, one can also define the associated equivalence relation, $\sim$,

[20] See Beja and Gilboa (1992) for necessary and sufficient conditions for the existence of such a representation, as well as an alternative for which strict preference is represented by a weak inequality, and indifference $\smile$ – by a strict inequality.

as follows: for every x, y, $x \sim y$ if and only if

$$\begin{aligned} \forall z, \qquad x &\succ z \Leftrightarrow y \succ z \\ &and \\ \forall z, \qquad z &\succ x \Leftrightarrow z \succ y \end{aligned}$$

Naturally, $x \sim y$ implies $x \smile y$, but the converse is not generally true. Indeed, $\sim$ is an equivalence relation, and, given a representation of $\succ$, (u, δ), one may assume that it also satisfies

$$x \sim y \qquad iff \qquad u(x) = u(y) \tag{7}$$

Under some richness conditions, this will follow from (5). In particular, this is the case if the range of u is the entire real line (as is assumed in this paper).[21]

[21] Note, however, that this is not always the case: if, for example, $\succ$ is empty, one can still represent it by a non-constant u as long as its range is contained in a δ-long interval.

Bibliography

Algom, D. (2001), "Psychophysics", in the *Encyclopedia of Cognitive Science*, Nature Publishing Group (Macmillan), London. 800-805.

Beja, A. and I. Gilboa (1992), "Numerical Representations of Imperfectly Ordered Preferences (A Unified Geometric Exposition)", *Journal of Mathematical Psychology*, **36**: 426-449.

Bentham, J. (1780), Introduction to the Principles of Morals and Legislation.

Blackorby, C., D. Primont, and R. R. Russell (1978), *Duality, Separability, and Functional Structure: Theory and Applications*, North Holland, 1978.

Borgers, T., Krahmer, D., and Strausz, R. (2015). *An introduction to the theory of mechanism design.* Oxford University Press, USA.

Dhillon, A. and J.-F. Mertens (1999), "Relative Utilitarianism", *Econometrica*, **67**: 471-498.

Edgeworth, F. Y. (1881). *Mathematical Psychics: An Essay on the Application of Mathematics to the Moral Sciences.* Kegan Paul, London. (Reprinted 1967, Augustus M. Kelley, New York.)

Fechner, G. T. (1860), *Elemente der Psychophysik*, 2 vol. (Elements of Psychophysics).

Fleurbaey, M. and Ph. Mongin (2016), "The Utilitarian Relevance of the Aggregation Theorem", *American Economic Journal: Microeconomics*, **8**: 289-306.

Fleurbaey, M. and S. Zuber (201), "Fair Utilitarianism", *mimeo.*

Fleming, M. (1952), "A Cardinal Concept of Welfare", *The Quarterly Journal of Economics*, **66**: 366-384.

Gilboa, I. and R. Lapson (1995), "Aggregation of Semi-Orders: Intransitive Indifference Makes a Difference", *Economic Theory*, **5**: 109-126. Working paper version, 1990, Northwestern University Discussion Paper No. 870.

Gul, F. and W. Pesendorfer (2006): "Random Expected Utility", *Econometrica*, **74**: 121–146.

Harsanyi, J. (1953), "Cardinal Utility in Welfare Economics and in the Theory of Risk-Taking", *Journal of Political Economy,* **61**: 434-435.

Harsanyi, J. (1955), "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility", *Journal of Political Economy,* **63**: 309-321.

Jevons, W. S. (1866), "Brief Account of a General Mathematical Theory of Political Economy", *Journal of the Royal Statistical Society*, **XXIX**: 282-287.

Luce, R. D. (1956), "Semiorders and a Theory of Utility Discrimination", *Econometrica,* **24**: 178-191.

Luce, R. D. (1958): "A Probabilistic Theory of Utility", *Econometrica*, **26**: 193–224.

Mas-Colell, A., Whinston, M. D., & Green, J. R. (1995). *Microeconomic theory (Vol. 1).* New York: Oxford university press.

Menger, C. (1871), *Grundsatze der Volkswirthschaftslehre.* Vienna. Translated as *Principles of Economics,* 1985.

Myles, G. D. (2008), *Public Economics.* Cambridge University Press.

Ng, Y.-K. (1975) "Bentham or Bergson? Finite Sensibility, Utility Functions, and Social Welfare Functions", *Review of Economic Studies*, **42**: 545-570.

Roberts, F. S. (1971), "Homogeneous Families of Semiorders and the Theory of Probabilistic Consistency", *Journal of Mathematical Psychology*, **8**: 248-263.

Roberts K. W. S. (1980), "Interpersonal Comparability and Social Choice Theory", *Review of Economic Studies*, **47**: 421-439.

Rubinstein, A. (1988), "Similarity and Decision-Making Under Risk", *Journal of Economic Theory*, **46**: 145-153.

Saez, E. and S. Stantcheva (2016), "Generalized social marginal welfare weights for optimal tax theory." *The American Economic Review*, **106.1**: 24-45.

Segal, U. (2000), "Let's Agree that All Dictatorships Are Equally Bad", *The Journal of Political Economy*, **108**: 569-589.

Sen, A. K. (1977), "On Weights and Measures: Informational Constraints in Social Welfare Analysis", *Econometrica*, **45**: 1539-1572.

Stanton, A. (2011), "Negishi Welfare Weights in Integrated Assessment Models: The Mathematics of Global Inequality", *Climactic Change*, **107**: 417-432.

Walras, M.-E.-L. (1874), *Elements d'economie politique pure.*

Weber, E. H. (1834), *De Tactu* ("Concerning Touch").