Weighted Utilitarianism, Edgeworth, and the Market*

Rossella Argenziano† and Itzhak Gilboa‡

June 2016

Abstract

Social choice involves trading off agents' utilities, and often people share strong moral intuitions about the way such trade-offs should be addressed. Yet, the textbook derivation of a utility function suggests that utility is only ordinal, and that interpersonal comparisons of utility are meaningless. We argue that comparisons of utility differences, and the use of social welfare functions, are rendered meaningful by information contained in real choice data that the textbook model ignores. In particular, human perception has limited and measurable accuracy. Just noticeable differences (jnd's) single out utility functions that are almost unique and can be used to make interpersonal comparisons of utility. Next, we show that a very weak monotonicity condition on society's preferences necessitates weighted utilitarianism, where the weights are the inverse of the individual jnd's. This was first suggested by Edgeworth, and reflects common moral intuitions. Finally, we use the language of weighted utilitarianism to enrich the debate over free markets and their normative appeal. Competitive markets compute an allocation that maximizes a weighted utilitarian function, with Negishi weights reflecting a "one dollar one vote" principle. We contrast this with the "one person one vote" principle reflected by the Edgeworth weights.

*We thank Paul Milgrom, Uzi Segal, and David Schmeidler for comments and discussions. We are particularly grateful to Luigi Balletta and Ludovic Renou for comments on earlier versions of this paper. 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

Frank and Ted are professors of economics, working in the same department. Frank is disabled and uses a wheelchair to move about campus. Ted can walk. Both Frank and Ted, who drive to campus, would like to have a reserved parking spot, preferably just by the department building. Unfortunately, only one such spot is available, so that one of them could get it and the other would have to roam about campus to look for a spot and then get to his office. Who should get it?

Pareto dominance clearly does not rank the two alternatives. Frank might say that he should get it, because it is harder for him to move around campus than it is for Ted. It stands to reason that Ted would agree. But he might still say, "OK, the parking is yours. But now we're both about to teach an intermediate microeconomics class where we say that interpersonal comparisons of utility have no scientific basis. How do we make these comparisons in practice, and why do we teach our students that they don't make sense?"

Indeed, microeconomic textbooks typically warn the student that utility functions are but mathematical artifacts used to represent preferences. They are shown to be only ordinal and it is emphasized that no particular meaning should be attached to their values or to differences thereof. This realization, 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, it is 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. For that reason, many economists would restrict normative questions to Pareto optimality, and shy away from trading off one individual's utility against another's.

At the same time, economic reality forces societies and governments to make choices that imply such trade-offs, and therefore cannot be dictated by Pareto domination alone. In some of these cases, a reference is made to social welfare functions as guiding principles for presumably ethical choices. In particular, utilitarianism (Bentham, 1780) is often invoked in contexts such as climate change discussions (see Stanton, 2011), mechanism design (see Borgers, 2015) and others. Evidently, such a social welfare function can capture our moral intuition in the example above, and assign a higher aggregated value to the alternative in which Frank gets the parking spot. But it would appear to be divorced from the purely ordinal notion of utility that is used to represent individuals' preferences.

Our first contribution is to address this uneasy conflict between the textbook foundations of neoclassical utility functions and their use for interpersonal comparisons in social welfare functions. We observe that this tension is an artifact of unrealistic assumptions of the standard model regarding what is and what isn't observable in consumer choice. Choice data that are actually available are much richer than the neoclassical consumer model admits. Moreover, naturally available data contain enough information to allow for comparison of utility differences within *and across* individuals, and may thereby endow a social welfare function with empirical foundations.

There are many types of choice data that psychology considers as valid observations and that are ignored by the neoclassical model. These include self-reports, response times, f-MRI scans, and others. In principle, one can take each of these sources of information as a basis for a cardinal measure of utility and employ it to make interpersonal comparisons of utility.

Of the various types of data that exist in actual choices, we focus here on one and investigate its implications for social choice. It is based on the limited accuracy of human perception: 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. Weber's Law (1834) suggests that the jnd is proportional to the base rate of the stimulus. Thus, a jnd emerges whenever there is a certain fixed increase on a logarithmic scale of the stimulus level.

Luce (1956) pointed out that, due to the fact that jnd's are positive, an indifference relation will typically not be transitive, and that there is a need for a general theory of decision making that can capture such intransitivity of indifference. He axiomatically defined binary relations, dubbed *semi-orders*, and showed that (in the finite case) they can be represented by a utility function and a positive threshold, such that strict preference emerges for one alternative over another only if the utility difference between them is above the threshold. In a case of a single good, a decision maker who prefers more to less and obeys Weber's Law would have semi-ordered preferences with the logarithm of the quantity serving as the utility function. Thus, Luce's axioms pointed out to the underlying structure of Weber-type preferences, and suggested to extend this structure to a general model of utility maximization. It turns out that utility functions that represent semi-orders are not ordinal; they are almost unique, where the utility threshold – the utility just-noticeable-difference – serves as a unit of measurement on the utility scale.

Probabilities of choice, which are used to define jnd's, differ from other sources of data, such as response times, self-reports, MRI scans and so forth, which the standard microeconomic model also ignores. The difference is that the fact that jnd's are positive has an implication, namely that indifference is not transitive, which stands in direct contrast to an explicit assumption of the model. This fact, coupled with Luce's work, is the reason we choose jnd's as our starting point to enrich the neoclassical model.

The main contribution of this paper is to model the aggregation of semi-ordered preferences. Once the notion of jnd's is introduced into the model of consumer choice, one obtains an almost-unique utility function for each individual, and these functions can be used as arguments of any social welfare function. However, our main result shows that, if individuals and society have semi-ordered preferences, a very weak condition implies a specific social welfare function: society's utility should be the weighted utilitarian aggregation of the individuals'. The condition we use is the equivalent of monotonicity in the case of weak orders. It turns out that such a weak condition suffices to derive a social welfare function that is separably additive, and to determine the individuals' relative weights in the aggregator. Further, these weights are precisely those suggested by Edgeworth (1881) in order to operationalize utilitarianism: he proposed that weights be so determined as to equate the jnd's across individuals.

We thus conclude that the conceptual gap between the ordinal utility functions of microeconomics textbooks and economic practice, which may resort to utilitarianism, is not as large as it may seem. Within the standard model, where preferences are captured by weak orders (with transitive indifferences), interpersonal comparisons of utility appear to be divorced from empirical content. But in a slightly more realistic model, admitting more data and the entailed complexity of choice, the gap can be bridged. Actual data allow one to pinpoint utility functions for individuals, and suggest a way of aggregating them.

We do not suggest that utility functions of all individuals in a society be measured so as to maximize their weighted sum. Such a proposition would raise a variety of practical problems. However, our model shows that such a measurement could be carried out in principle, and that the notion of social welfare functions in general, and weighted utilitarianism in particular, are not divorced from the empirical foundations of utility theory. In practice, one might use weighted utilitarianism as a guideline in policy questions using utility functions and weights that depend only on individuals' health and age, as a proxy for their true jnd's. Importantly, this would suffice to justify the allocation of parking spots to disabled people as in the example above.

We conclude the paper by observing that the language of weighted utilitarianism can enrich the debate over free markets and their normative appeal. We state two rather obvious results about weighted utilitarianism, corresponding to the welfare theorems of neoclassical economics. Specifically, competitive markets can be viewed as "computing" an allocation that is a maximizer of a weighted utilitarian function.

While these results can be stated for any specification of utility functions for the individuals, the semi-orders perspective allows us to interpret utilitarian weights in a meaningful way. In particular, it is shown that the weights that correspond to the market solution reflect a "one dollar one vote" principle, as opposed to Edgeworth's "one person one vote" principle. It is argued that this perspective allows us to capture some of the positive as well as negative moral sentiments that competitive markets evoke.

This paper is organized as follows. Section 2 presents the notion of just noticeable differences and discusses their implications to the measurement of utility. Section 3 explains Edgeworth's proposal and presents the main result of this paper, deriving this proposal from a weak condition named "Consistency". Section 4 proceeds to discuss the utilitarian welfare observations. Finally, Section 5 concludes.

2 Measurability of Utility

2.1 Semi-Orders

The textbook microeconomic model suggests that all that is observable are consumer choices, typically modeled as a complete and transitive binary relation over alternatives. It is a mathematical fact that a utility function representing such a relation can be replaced by any (strictly increasing) monotone transformation thereof without changing the implied preferences. However, it is wrong to assume that choices are given only by binary preference relations that are complete and (always) transitive. In particular, real choices systematically deviate from transitivity of indifferences.

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

$$\log (S + \Delta S) - \log (S) > \delta \equiv \log (C) > 0. \tag{1}$$

This law is considered a rather good first approximation and it appears in most introductory psychology textbooks.[1]

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,

[1] 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).

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 the ability to discern n grains from $(n+1)$ grains of sugar in an (otherwise identical) cup of coffee goes down in n. Thus, starting with a small enough grain, 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. The amount of food we consume, the temperature of our house, the duration of our vacation – almost all our experiences involve quantities that can be measured with greater precision than our perception can discern. Luce therefore 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:

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,

$$x \succ y \qquad iff \qquad u(x) - u(y) > \delta \tag{2}$$

$$x \smile y \qquad iff \qquad |u(x) - u(y)| \leq \delta \tag{3}$$

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.[2]

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 (2) (and the implied (3)) 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, $\sim$, as follows: for every x, y, $x \sim y$ if and only if

$$\begin{array}{rrcl} \forall z, & x & \succ & z \Leftrightarrow y \succ z \\ & & & and \\ \forall z, & z & \succ & x \Leftrightarrow z \succ y \end{array}$$

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

$$x \sim y \qquad iff \qquad u(x) = u(y) \tag{4}$$

[2] 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.

Under some richness conditions, this will follow from (2). In particular, this is the case if the range of u is the entire real line (as will be assumed in the sequel).[3]

Luce referred to the relation $\smile$ as "indifference", and the literature typically discusses "intransitive indifference". One might argue, though, that true indifference isn't captured by $\smile$, but by $\sim$. For example, a person who prefers as little sugar in the coffee isn't truly "indifferent" between close quantities, even if she can't tell them apart. This inability to discern close magnitudes is due to a problem of measurement, but the true preferences of the individual are given by a classical weak order. Indeed, if we were to equip her with a kitchen scale, thereby improving her measurement accuracy, her true, perfectly transitive preference would emerge.

We have no strong opinions on the proper usage of the word "indifference". Following common practice, we will use it to refer to the relation $\smile$, without taking a stand on the nature of "true indifference". Our main point is that actual preferences contain more information than the standard model admits. To prove this point it suffices to show that preferences that rely on perception, without the aid of scientific measurement devices, offer implicit comparisons of *differences* between utility levels: those that are above vs. those that are below a certain threshold. This distinction between "large" and "small" differences provides a scale that allows one to measure differences in utility between any two alternatives, basically, by asking how many "large" differences one can fit between the utility levels. And if Luce's axioms hold (as implied by Weber's Law in the case of perception), one obtains a utility function u and a number $\delta > 0$ that satisfy (2, 3), and (4).

[3]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.

2.2 Uniqueness of Utility

It is easy to see that if (2) is the notion of "representation of preferences" one has in mind, the utility function u used in it is not only ordinal. Assume that the alternatives are points in a connected space such as $\mathbb{R}^l$ and that the utility function u is continuous. In this context, one may indeed consider arbitrary increasing transformations of the u that retain differences under δ, and get other functions that also represent preferences as in (2). Specifically, for any strictly increasing function

$$f : \mathbb{R} \rightarrow \mathbb{R}$$

if, for every $\alpha, \beta \in \mathbb{R}$,

$$|\alpha - \beta| \leq \delta \qquad iff \qquad |f(\alpha) - f(\beta)| \leq \delta$$

then

$$v = f(u)$$

represents $\succ$ as in (2) if and only if u does.

Thus, there is a great deal of freedom in selecting the function u "in the small". Indeed, the function f above can be any arbitrary strictly increasing function over the $[0, \delta]$ interval, as long as

$$f(\delta) - f(0) = \delta.$$

But the number of "δ-steps" between two alternatives has to be respected by any function that represents preferences, whether measured on the original u scale or on the transformed v scale.

And the number of just-noticeable-difference (δ) steps between alternatives can provide a measure of the intensity of preferences. For example, if we consider three alternatives $x \succ y \succ z$ such that

$$4\delta < u(x) - u(y) \leq 5\delta$$

but

$$\delta < u(y) - u(z) \leq 2\delta$$

it is meaningful to say that "x is better than y *by more than* y is better than z".

Moreover, one can provide empirical meaning to claims such as "the marginal utility of money is decreasing". 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". Importantly, the above is based on observable data.

2.3 JND's and Utilitarianism

What is the upshot of this discussion? There are some popular claims in welfare economics which, we hold, would qualitatively change when semi-orders are taken into account. The claims we take issue with are:

1. Utility is "only ordinal".

2. There is no meaningful way to make interpersonal comparisons of utility.

3. Therefore, utilitarianism cannot be operationalized.

We have devoted subsection 2.2 to show that (1) does not hold in the presence of semi-orders. That is, the utility function is "much more unique" than standard consumer theory would have us believe. However, even if we had a cardinal utility for each agent (as implied, say, by preferences over vNM lotteries), claim (2) might independently hold. We now wish to make a bolder claim, namely that the jnd scales offer a way to make interpersonal comparisons of utility, based on equating jnd's across individuals.

This claim being rather bold, we split it into two: first, we make the trivial observation that jnd's do offer an empirically meaningful way to compare utilities. Then, we will try to make the more challenging step, arguing that these comparisons of utilities make sense. At this point we ask the reader to accept that the existence of jnd's in actual data invalidate both claims (1) and (2), and that their conclusion consequently does not hold. Specifically, jnd's are in-principle observable, and they allow us to find, for each agent an almost-unique utility function, and, further, a way to compare these.

We now turn to the bigger challenge, of convincing the reader that the jnd scales do capture some intuition, of some "moral sentiments" that they may share. We devote the next section to this end.

3 Edgeworth's Version of Utilitarianism

3.1 Edgeworth's Ethical Solution

Realizing that utility functions representing semi-orders via just noticeable differences are almost unique, one can revisit the question of utilitarianism and ask whether these almost-unique functions can be assigned reasonable weights in an additive social welfare function (SWF). Edgeworth (1881, p. 60) wrote,

> "Just perceivable increments of pleasure, of all pleasures for all persons, are equateable."

That is, there is a natural set of weights that are ethically appealing: weights that equate the just noticeable differences across individuals. To be more precise, assume that $x, y, ...$ denote social alternatives such as consumption allocations, and that each individual i has semi-ordered preferences $\succ_i$ over them, represented by

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

The representation of a preference order $\succ_i$ by (u_i, δ_i) can be replaced by $(au_i, a\delta_i)$ for any $a > 0$. Without loss of generality we may assume that $\delta_i = 1$, that is, replace (u_i, δ_i) by $(\frac{1}{\delta_i}u_i, 1)$. We agree with Edgeworth that, with this normalization, it is natural to assign to all individuals the same weight in the weighted utilitarian function. We will try to convince the reader of this claim by a few examples in this sub-section, and by a theorem in the next one.

Consider the following example. An old man is carrying a heavy bag, and a healthy, athletic youngster is walking next to him cheerfully and leisurely. We would probably feel that it would be nice should the youngster offer to carry the bag – much more than we would find it acceptable if the situation were reversed. Asked why, a person might say that for the youngster it is "no big deal" to carry the bag, whereas the task is very difficult for the old man. We argue that jnd's calculus offers one way to capture this intuition: the "no big deal" argument can be mapped to "very few jnd's" whereas the "very difficult" – to "many jnd's". Observe that in this section we are not trying to convince youngsters to help the elderly. We only argue that the jnd version of utilitarianism is not a bad mathematical model for the "no big deal argument".

Further, we claim that a re-allocation of property rights to match such utilitarian solutions does indeed occur in the society we live in. Consider Frank and Ted we started out with. Frank, being confined to a wheel chair, could be viewed as having a small jnd than does Ted. For example, assume that apart from the privileged parking spot all others require climbing three steps on the way to the department. Ted, being healthy and fit, might be walking to class without even noticing the steps. This will not be the case for Frank, who has to maneuver the wheel chair around or ask for help in climbing the stairs. Thus, according to jnd calculations, we can conclude that it is harder for Frank to climb the stairs than it is for Ted.[4] Thus, Edgeworth's solution would suggest that Frank get a higher weight in the social welfare function, and therefore that it is optimal that he would get the preferred parking lot rather than would Ted. Treating people equally, after having taking into account their sensitivities, is akin to "one person one vote", and will be referred to as *the ethical solution*.

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 vNM 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.

3.2 A Formal Derivation

We now turn to derive Edgeworth's version of utilitarianism from a simple condition. 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).[5]

[4] Note that Ted could try to argue that this is not the case, and that, despite his health, climbing three steps involves a huge mental cost to him. As mentioned in the Discussion, we do not delve into the messy issue of manipulability of reported jnd's here. We note, however, that in practice, while Ted could pretend to be disabled as well, if he doesn't do so successfully, society would not grant him the property rights bestowed upon those it judges to be less fortunate.

[5] 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.

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 $\mathbb{R}^l_{++}$ that is represented by (u_i, δ_i) as follows: for every $x_i, y_i \in \mathbb{R}^l_{++}$,

$$\begin{aligned} x_i \succ_i y_i \quad & iff \quad u_i(x_i) - u_i(y_i) > \delta_i \\ x_i \smile_i y_i \quad & iff \quad |u_i(x_i) - u_i(y_i)| \leq \delta_i \end{aligned} \tag{5}$$

We assume that u_i is weakly monotone and concave, and that $\delta_i > 0$. The main result of this section can be significantly generalized in terms of the assumptions on the domains of preferences and their structure. However, we remain as close as possible to the standard general equilibrium model. Observe that, because the domain $\mathbb{R}^l_{++}$ is open, the concave utility u_i has to be continuous.

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 (5) implies that u_i is unbounded, and its continuity implies that $range\,(u_i) = \mathbb{R}$.

An allocation is an assignment of bundles to individuals,

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

We assume that society has semi-ordered preferences $\succ_0$ on the set of allocations $\left(\mathbb{R}^l_{++}\right)^n$ that is represented by (u_0, δ_0) with $\delta_0 > 0$.[6] Without loss of generality we assume that $\delta_0 = 1$. Thus, $u_0 : \left(\mathbb{R}^l_{++}\right)^n \to \mathbb{R}$ is such that, for every $x, y \in \left(\mathbb{R}^l_{++}\right)^n$,

$$\begin{aligned} x \succ_0 y \quad & iff \quad u_0(x) - u_0(y) > 1 \\ x \smile_0 y \quad & iff \quad |u_0(x) - u_0(y)| \leq 1 \end{aligned} \tag{6}$$

[6] One might wonder whether society's preferences should be given by a semi-order as are the individuals'. We find it more coherent to have all preference relations being of the same family. However, if one were to insist that society have only a weak order, one can re-phrase our Consistency axiom below to obtain a similar result.

We similarly assume that u_0 is concave, hence 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 replacing the i-th component of z, z_i, by x_i. The main assumption we use is

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

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

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.

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 should simply be those of the individual. 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. However, Consistency also requires that society not be more sensitive than the individual herself. If individual i cannot tell the difference between x_i and y_i, it is assumed that the difference between the two is immaterial to society as well. This assumption appears rather intuitive if we interpret $x \smile_0 y$ as "too close to be worth worrying about". 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.[7]

Importantly, Consistency does *not* require that society agree 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.

[7]If, however, one drops the "only if" part of the axiom, or uses only a standard weak order for society, a similar result can be obtained with an additional condition that says 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 individual.

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. Still, Consistent does not imply that society should prefer this donation scheme. Indeed, requiring this implication would result in a stronger assumption that leads to intransitivities (as one can change the happy recipient of the individually-negligible donations and generate cycles of strict societal preferences).

Rather, Consistency is restricted to the case that no individual $j \neq i$ is affected at all, whether he 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).

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* $x \in A$,

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

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. Thus, Edgeworth's suggestion, which we find rather intuitive in its own right, can be further supported by a relatively simple and innocuous condition.

Given the implications of the Consistency axiom, one might wonder, is it perhaps too strong? Does it look innocuous while, in fact, assuming more than is reasonable? Suppose, for example, that instead of n individuals we discuss n goods, and the preferences are those of a single consumer. The axiom would then imply that the consumer has a separably additive utility function. Is that not too restrictive?

Indeed, if we were discussing a descriptive model of consumer behavior, Consistency might not be a reasonable assumption to adopt. A consumer's jnd for one good might well depend on the quantities of other goods. However, when different individuals are concerned, it seems plausible, and certainly normatively appealing, to assume that each individual's jnd is defined on her bundle alone, and use that jnd when judging allocations that only differ in that individual's bundle.

4 The Utilitarian Welfare Observations

4.1 The Goal

Strength of preferences can often be measured using one of the goods. For example, we can ask how long is one willing to wait to get a certain bundle, or, more simply, how much is willing to pay for it. In a sense, this is what a free market would do for us: it would allocate a good to the individual who's willing to pay more for it.

The difficulty with this way of measuring strength of preferences is that it depends on initial endowments. Considering Frank and Ted again, if Ted is much richer than Frank, he might be able to purchase the right for the coveted parking spot even though we would feel that Frank "needs it more".

More generally, we2 wish to contrast the set of weighted utilitarian solutions with competitive equilibria allocations. We adopt the model of the previous section and assume that the utility function u_i for individual i is normalized with a jnd of $\delta_i = 1$ (so that $\frac{1}{\delta_i} = 1$). Hence, Edgeworth's ethical solution, independently derived from Consistency in Theorem 1, corresponds to the maximization of a social welfare function

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

However, there are other weighted utilitarian welfare functions, that need not equate the individuals' weights. Indeed, once we used the jnd's to fix a utility function u_i for each individual, we can still define, for weights $\lambda = (\lambda_1, \ldots, \lambda_n)$ (with $\lambda_i > 0$), the λ-weighted utilitarian function to be

$$U_\lambda(x) = \sum_{i \leq n} \lambda_i u_i(x).$$

In the neoclassical model, where (u_i) represent transitive preferences (with transitive indifferences), it has been observed by Negishi (1960) that the set of maximizers of U_λ, for different λ, coincides with the set of allocations that can be obtained as competitive equilibria. Indeed, both are identical to the set of Pareto optimal allocations. When (u_i) represent semi-ordered preferences (with appropriate jnd's (δ_i)), a similar conclusion is to be expected. However, one has to define equilibria and Pareto efficiency for the claim to be well-defined. We do so in the next subsection, and proceed to discuss the conceptual issues in the following one.

4.2 The General Equilibrium Model

Let $e^j > 0$ ($j \leq l$) be the total quantity of good j. Consider an initial endowment $e \in \left(\mathbb{R}^l_{++}\right)^n$ with $\sum_i e_i^j = e^j$ (for every $j \leq l$) and the exchange economy defined by the utilities $(u_i)_i$ and e.

For every feasible allocation $x \in \left(\mathbb{R}^l_{++}\right)^n$ (that is, an allocation such that $\sum_i x_i^j \leq e^j$) let $u(x)$ be the utility profile defined by x.

First, we note the following.

Remark 1 The set

$$F \equiv \left\{ u(x) \middle| x \in \left(\mathbb{R}^l_{++}\right)^n, \sum_i x_i^j \leq e^j \right\}$$

is convex. A point $u \in F$ is on the Pareto frontier of F if and only if it is a maximizer of U_λ for some $\lambda >> 0$.

An *equilibrium* is defined as pair (p, x) such that $p \in \mathbb{R}^l_{++}$ is the price vector and $x \in \left(\mathbb{R}^l_{++}\right)^n$ is the corresponding feasible allocation, such that no agent can increase her u_i by more than her jnd δ_i within the budget set defined by her endowments and the prices. Formally, for all $i \leq n$ it is required that

$$u_i(x_i) \geq u_i(y_i) - \delta_i$$

for all $y_i \in \mathbb{R}^l_+$ such that

$$py_i \leq pe_i.$$

Notice that the feasibility constraint is defined without reference to the jnd's, as it is presumed to be a physical constraint on quantities bought in the market. When the agents run out of money, they stop shopping even if the were under the impression that they weren't consuming more than before. By contrast, the optimality constraint includes the jnd: in order to consciously choose to switch from bundle x_i to y_i, the agent needs to notice that the latter will be better than the former.

The existence of equilibria is immediate:[8] consider the standard economy where agents have transitive preferences defined by $(u_i)_i$ and endowments e_i. For any equilibrium p of this standard economy and any corresponding equilibrium allocation x, the pair (p, x) is an equilibrium of our economy.

[8] See also Jameson and Lau (1977).

Clearly, our definition allows for more equilibrium allocations, including also those that deviate from a standard equilibrium by less than noticeable differences. For simplicity, in the sequel we focus on the allocations that are defined by precise maximization of u_i for each agent.[9]

Next, we state two observations that bear a conceptual resemblance to the classical welfare theorems, with the weighted utilitarian criterion replacing Pareto optimality.

Observation 1 *The First Utilitarian Welfare Observation: Let* $x \in \left(\mathbb{R}^l_{++}\right)^n$ *be an allocation of a competitive equilibrium for an endowment* $e \in \left(\mathbb{R}^l_{++}\right)^n$*. Then there exists a set of weights* $\lambda >> 0$ *such that* $u(x)$ *is a maximizer of* U_λ*.*

Conversely,

Observation 2 *The Second Utilitarian Welfare Observation: Let there be given* $\lambda >> 0$ *and an allocation* $x \in \left(\mathbb{R}^l_{++}\right)^n$ *such that* $u(x)$ *is a maximizer of* U_λ*. Then there exists an endowment* $e \in \left(\mathbb{R}^l_{++}\right)^n$ *such that* x *is a competitive equilibrium allocation of the economy with endowment* e*.*

The immediate proofs of these observations are given in Appendix A for the sake of completeness.

4.3 Competitive Equilibria and Ethics

We have discussed Edgeworth's suggestion for the choice of welfare weights that equate jnd's, referred to as "the ethical solution". However, markets may not "compute" this solution. Which solutions do they compute?

Negishi (1960) pointed out that each competitive equilibrium maximizes a weighted utilitarian welfare function, and that the welfare weights are the inverse of marginal utilities of income. Consider the exchange economy above

[9] One can similarly extend the definitions to include economies with production.

with an aggregate amount of $e^j > 0$ of good $j \leq l$. A competitive equilibrium allocation x maximizes a weighted welfare function:

$$U_\lambda (x) = \sum_{i \leq n} \lambda_i u_i (x_i)$$

If x is an interior point, to maximize $U_\lambda (x)$ we need it to be the case that, for every $j \leq l$, and every $i, k \leq n$,

$$\frac{\frac{\partial u_i}{\partial x_i^j} (x_i)}{\frac{\partial u_k}{\partial x_k^j} (x_k)} = \frac{\lambda_k}{\lambda_i} \tag{7}$$

This first order condition has the following troubling property: the more one has at an equilibrium (of any given good, other things being equal), the lower is one's marginal utility, and the higher is the weight one would need to have in the utilitarian social welfare function in order to justify the equilibrium allocation as a weighted utilitarian solution.

We point out that Negishi's (1960) main motivation was to use the weights as a mathematical tool, used to prove existence of equilibria. Indeed, this mathematical technique has been used in many subsequent works without making any normative claims. (See Young, 2008, for a survey.) In some cases, these weights have been used for normative purposes, where the weights are considered to be the accepted status quo. For example, in climate change debates, it seems impractical to apply equal weights to all regions around the globe, as these would suggest an immediate transfer of wealth from rich to poor regions, independently of climate effects. Such a proposal might be appealing to some, but it is considered to be impractical and unrelated to the environmental debate. Hence, in such contexts the Negishi weights are sometimes adopted as an accepted starting point, used to determine the appropriate course of action *given* existing inequality. (See, for instance, Stanton, 2011.) Our application of the weights is much simpler: we only use them as another way to capture dislike of inequality in a static model.

For the sake of the argument, assume that each agent satisfies Consistency over goods, and that her perception of increments in each good follow Weber's Law as in (1). These assumptions readily imply[10] that each agents's preferences can be described by the (log-linear representation of) a Cobb-Douglas utility function, so that

$$u_i(x) = \sum_{j=1}^{l} \alpha_i^j \log\left(x_i^j\right).$$

Then

$$\frac{\partial u_i}{\partial x_i^j}(x_i) = \frac{\alpha_i^j}{x_i^j}$$

and condition (7) states that, for every $j \leq l$, and every $i, k \leq n$,

$$\frac{\lambda_i \alpha_i^j}{x_i^j} = \frac{\lambda_k \alpha_k^j}{x_k^j}.$$

Next assume that for at least one good the sensitivity of all individuals is identical. Specifically, suppose that we consider people who are similar in terms of the physiology, and good $j = 1$ represents a basic necessity, such as calorie intake or sleep. Alternatively, we may think of good 1 as representing the amount of money the individual saves for her children. If we can then assume that α_i^1 is independent of i, we get

$$\frac{x_k^1}{x_i^1} = \frac{\lambda_k}{\lambda_i}.$$

Or, without loss of generality, $\lambda_i = x_i^1$. If x_i^1 denotes money saved for the future, we find that the market mechanism maximizes a utilitarian welfare function in which the weight of each individual is her wealth: one dollar, one vote.

Clearly, it is not the market mechanism per se that is the source of this apparent inequality. Free trade only seeks Pareto improvements over the

[10] See Appendix B for details.

initial endowments, and these are to be blamed for inequality. The implicit claim made here against competitive equilibria is not that they generate inequality, but that they accept it. To illustrate, we can imagine an economy as above with only one good ($l = 1$), say "money". Clearly, any allocation is Pareto optimal, there is no room for trade, no prices to speak of, and any allocation is also an equilibrium allocation in this trivial sense. It is still true that any such allocation maximizes a weighted utilitarian function, and that "one dollar one vote" applies to the implied weights of such a function.

We believe that the utilitarian analysis above captures some of the ethical reactions to competitive markets: on the one hand, they guarantee Pareto optimality, and that would be considered normatively appealing by most. On the other hand, in order to explain which Pareto optimal allocation got selected by the market mechanism, one has to assume that the rich are weightier than the poor.

5 Discussion

5.1 Manipulability

Another notorious problem with the implementation of weighted utilitarianism is manipulability: how will we find individuals' "true" utility functions? Will they not have an incentive to misrepresent their choices in order to obtain a higher weight in the utilitarian social welfare function?

We observe that

– If we adopt the inverse of the jnd as a person's weight in the SWF, manipulation isn't easy to accomplish. One may pretend to be less sensitive than one actually is, but this would decrease one's weight in the function, not increase it. And in the absence of the ability to discern small differences, one cannot get one's weight to be higher than it "should" be.

– A common approach to deal with such problems is to ignore individuals' stated or measured utility function, and to use instead a utility function that is ascribed to them by a social planner. This is, arguably, what societies do when they provide welfare to the poor (and not to the rich), select progressive tax schedules, provide medical treatment to the sick, and so forth. Along these lines, the analysis above may help us conceptualize the social justice problem without necessarily measuring individual utility functions.

5.2 Endogenous Just Noticeable Differences

It is important to note that the Edgeworth suggestion should be understood in the context of a single-period model. In a multiple-period model, one should take into account the possibility that jnd's change, in particular as a result of education. Consider the following example. 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. Is this fair?

For many readers, the answer would be negative. The connoisseur most likely became one via experiences. Thus, she is a person who has had the good fortune to enjoy excellent wines and develop her taste. The layperson probably never had the chance to do so. It seems unfair to reinforce this inequality by allotting the good wine to the expert, leaving the layperson in his ignorance. The layperson can also learn to appreciate good wine, and it seems more fair to give them the chance to do so.

However, this intuition relies on the fact that tastes, and, in particular, jnd's change as a function of consumption. To deal with this problem, we would need at least two periods, with the possibility of changing tastes and uncertainty about these. In such an extended model the utilitarian calculation would not be that straightforward. Indeed, taking into account future agents' jnd's, one may argue that the layperson should be allotted the better wine so that her future selves would become more discerning, adding more jnd's, as it were, to the social welfare function. We leave the construction of such a dynamic model for future research.

5.3 von Neumann and Morgenstern Utilities

A common way to pin down a utility function for an individual involves extending the set of choices from consumption bundles to lotteries over such bundles. von Neumann and Morgenstern's (1944) theorem provides foundations for expected utility maximization, under which the utility function is unique only up to positive affine transformations. This degree of uniqueness is quite impressive, comparable to the uniqueness of the measurement of temperature. Still, the arbitrariness in determining the unit of measurement suffices to pose a problem for utilitarianism. Harsanyi (1953, 1955) and others attempted to deal with this arbitrariness by making some normalization assumptions, such as setting all utilities to a range of a given interval (see Dhillon and Mertens, 1999).

We find that this approach may be unsatisfactory for two reasons. First, it is not clear that the utility that describes behavior under risk should be the basis for resolution of ethical issues. For example, assume that two individuals $i = 1, 2$ have to share a loaf of bread, and we normalize their utilities so that $u_i(x) = x$ for both individuals ($i = 1, 2$) and for the extreme quantities, $x = 0, 1$. Assume that individual 1 is risk neutral, so that $u_1(x) = x$ for all $x \in [0, 1]$ while individual 2 is risk averse. The (unweighted) utilitarian solution would endow individual 2 with a smaller portion of the loaf than individual 1. This may indeed be a result of negotiation between them, where risk aversion can be a weakness in bargaining (see Kihlstrom, Roth, and Schmeidler, 1981). But many would doubt whether it is fair or ethical to penalize individuals for their risk aversion.

Second, the normalization of utilities of all individuals to a certain range rules out the possibility than one individual is perceived to be generally more sensitive than another. In our leading example, we might want to argue that Frank, suffering disability, can experience larger utility differences than can Ted, who isn't. If we set both of their utility functions to have a minimum at 0 and maximum at 1 we will end up with a model that does not capture the intuitive notion that Frank is more sensitive than Ted.

Appendix A: Proofs

Proof of Theorem 1:

First, assume Consistency.

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

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 $range\left(u_0\left(z_{-i}, \cdot\right)\right)$ is unbounded (from below and from above). Given that u_0 is concave (on an open set), hence continuous, its range is also convex, and $range\left(u_0\left(z_{-i}, \cdot\right)\right) = \mathbb{R}$ 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

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

so that

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

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 $u_i\left(y_i\right) > u_i\left(w_i\right) + \delta_i$, which implies $u_i\left(x_i\right) > u_i\left(w_i\right) + \delta_i$, 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 $u_i\left(x_i\right) > u_i\left(y_i\right)$, then $u_0\left(z_{-i}, x_i\right) > u_0\left(z_{-i}, y_i\right)$.

Proof: Assume that z, i, x_i, y_i are given with $u_i(x_i) > u_i(y_i)$. As $range(u_i) = \mathbb{R}$ 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 $u_0(z_{-i}, y_i) \le u_0(z_{-i}, w_i)+1$. 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 $i \le n$, $u_i(x_i) \ge u_i(y_i)$, then $u_0(x) \ge u_0(y)$.

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 $i \le n$, $u_i(x_i) = u_i(y_i)$, 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.

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 $k \ge 1$ such that $\alpha^k \to_{k\to\infty} \alpha$ 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 $range(u_i) = \mathbb{R}$ 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

$u_i\left(x_i^k\right) = \alpha_i^k$ and that $x_i^k \to_{k\to\infty} x_i$. If such a sequence existed, we would have $x^k = \left(x_1^k, ..., x_n^k\right) \to_{k\to\infty} x$ while

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

for every k, contradicting the continuity of u_0.

Consider, then $i \le n$ and $k \ge 1$. Let

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

As $range\,(u_i) = \mathbb{R}$, $A_i^k \neq \varnothing$. 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 $\alpha_i^k \to_{k\to\infty} \alpha_i$, 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 $x_i^k \to_{k\to\infty} x_i$ 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 $i \le n$,

$$g\left(\alpha + \delta_i 1_i\right) = g(\alpha) + 1$$

(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

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

By monotonicity of g, this can only hold if

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

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 $u_i\left(w_i\right) = u_i\left(x_i\right) - \delta_i > u_i\left(y_i\right)$ and show that $u_0\left((z_{-i}, w_i)\right) = u_0\left((z_{-i}, x_i)\right) - 1$ while $u_0\left((z_{-i}, w_i)\right) > u_0\left((z_{-i}, y_i)\right)$ so that $u_0\left((z_{-i}, x_i)\right) - 1 > u_0\left((z_{-i}, y_i)\right)$ and $(z_{-i}, x_i) \succ_0 (z_{-i}, y_i)$ follows. □□□

Proof of Remark 1:

The set F is convex because the utility functions are concave (and free disposal is allowed). Because the utility functions are strictly monotone and we consider only the interior of the feasible allocations, the supporting hyperplanes would not resort to zero coefficient, and the conclusion follows.[11] □

Proof of Observation 1:

Let there be given a competitive equilibrium allocation x for a strictly positive endowment $e \in \left(\mathbb{R}^l_{++}\right)^n$. Given the (classical) first welfare theorem, $x \in \left(\mathbb{R}^l_{++}\right)^n$ is Pareto optimal (in the standard sense), and $u\left(x\right)$ is a maximal point in F. Using Remark 1, x is a maximizer of U_λ for some $\lambda >> 0$. □

Proof of Observation 2:

Let $\lambda, x \in \left(\mathbb{R}^l_{++}\right)^n$ be given. Because $u\left(x\right)$ is a maximizer of U_λ (and λ is strictly positive), x is Pareto optimal (as stated in Remark 1). The (classical) second welfare theorem guarantees that x is an equilibrium allocation for the economy defined by $e = x$. □

[11] The Proposition in Yaari (1981, p.7), makes a similar observation, allowing for a closed domain and zero coefficients, but defining Pareto in the strict sense.

Appendix B: Consistency for Consumer Choice and Cobb-Douglas Preferences

We illustrate the implications of the axiom for a single agent's preference over consumption bundles. Suppose that we consider l different product categories such as food, entertainment, housing, and so forth. Consistency would then imply that the agent's preferences over bundles $x = \left(x^1, ..., x^l\right)$ can be represented by

$$u(x) = \sum_{j=1}^{l} \alpha^j v_j \left(x^j\right) \tag{8}$$

(with $\alpha^j > 0$ for $j \leq l$) over each jnd-grid. Further, assume that, for each category j, the relevant jnd is determined by Weber's Law as in (1). Thus, for each product j, if we vary the quantity x^j and seek the jnd, we should expect to find that it is proportional to x^j, and that an increase Δx^j will be noticeable iff

$$\log\left(x^j + \Delta x^j\right) - \log\left(x^j\right) > \delta^j. \tag{9}$$

In other words, the functions v_j are multiples of the logarithmic function. Combining (8) and (9) we get

$$u(x) = \sum_{j=1}^{l} \alpha^j \log\left(x^j\right) = \sum_{j=1}^{l} \frac{1}{\delta_j} \log\left(x^j\right) \tag{10}$$

which is the logarithmic representation of the widely used Cobb-Douglas functions.[12]

To conclude, we propose Consistency as a normative axiom for aggregating over different individuals' preferences, not as a descriptive one for aggregating over goods in a bundle. However, if we were to follow its logic we

[12] Clearly, with positive jnd's, the function (10) is no longer equivalent to

$$w(x) = \prod_{j=1}^{l} \left(x^j\right)^{\alpha^j}.$$

get a psychophysical foundation for the most popular example of consumer preferences.

Bibliography

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

Arneson, R. (2000), "Luck, Egalitarianism, and Prioritarianism", *Ethics,* **110**: 339-349.

Arrow, K. J. (1973), "Some Ordinalist-Utilitarian Notes on Rawls's Theory of Justice", *Journal of Philosophy,* **70**: 245-263.

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.

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

Edgeworth, F. Y. (1881). *Mathematical Physics.* Kegan Paul, London. (Reprinted 1967, Augustus M. Kelley, New York.)

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

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.

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.

Jamison, D. T. and L. J. Lau (1977), "The Nature of Equilibrium with Semiordered Preferences", *Econometrica*, **45**: 1595-1605.

Kihlstrom, R., A. E. Roth, and D. Schmeidler (1981), "Risk Aversion and Solutions to Nash's Bargaining Problem." In *Game Theory and Mathematical Economics*, edited by O. Moeschlin and D. Pallaschke, 65–71. Amsterdam: North-Holland Publishing Company.

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

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

Negishi, T. (1960), "Welfare Economics and Existence of an Equilibrium for a Competitive Economy", *Metroeconomica*, **12**: 92-97.

Rawls, J. (1971), *A Theory of Justice.* Belknap Press, Cambridge, MA.

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

Rubinstein, A. (2007), "Instinctive and Cognitive Reasoning: A Study of Response Times", *Economic Journal*, **117**: 1243-1259.

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

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

Stevens, S. S. (1975). *Psychophysics: Introduction to Its Perceptual, Neural, and Social Aspects.* John Wiley, New York.

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

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

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

Yaari, M. (1981), "Rawls, Edgeworth, Shapley, Nash: Theories of Distributive Justice Re-examined", *Journal of Economic Theory,* **24**: 1-39.

Young, W. (2008), "Negishi's Contributions to the Development of Economic Analysis: Research Programs and Outcomes", *International Journal of Economic Theory*, **4**: 151–165