IMBS TECHNICAL REPORT 18-01

Utility without Probability, Aggregation without Interpersonal Comparability: a Neo-Benthamite Approach*

Brian Skyrms† Louis Narens‡

January 30, 2018

Nature has placed mankind under the governance of two sovereign masters, pain and pleasure.

Bentham

Abstract

Most contemporary philosophical discussions of utility either follow the utility representations of von Neumann and Morgenstern that use probability and the expectation principle to measure individual utilities in a way that leaves zero point and unit undetermined, or assume that psychology can ultimately deliver utility measurement on an absolute scale. From the second point of view utilitarian aggregation is perfectly meaningful, and consequences of utilitarian ethics are straightforward. From the first utilitarian aggregation is perfectly meaningless. We develop a middle position, along the lines suggested by some of Bentham's ideas. The result is the measurement of individual utilities without the use of probability in a way that is weaker than that envisioned by Bentham but stronger than that given by von Neumann-Morgenstern. In favorable circumstances this allows meaningful aggregation, using a product or geometric mean, rather than the utilitarian sum. This has consequences for philosophical discussions of utility and utilitarianism.

*Research for this article was supported by Grant SMA-1416907 from National Science Foundation.

†Department of Logic and the Philosophy of Science, University of California, Irvine, Irvine CA92717 USA and Department of Philosophy, Stanford University, Stanford CA94305 USA.

‡Department of Cognitive Sciences University of California, Irvine, Irvine CA92717 USA.

1 Introduction

One of the many strands[1] in Bentham's thought about utility was the use of preference to resolve issues. Pleasure may be produced through many modalities. When confronted with the problem of multi-modal comparisons, Bentham said we must simply consult the subject's preferences:

> If, of two pleasures, a man, knowing what they are, would as lief enjoy one as the other, they must be reputed equal. (Bentham in Halévy v. 1 Appendix II p. 306)

Here we develop this strand of Bentham's thought. It does not lead exactly where Bentham wanted to go, but it does give something more than what we have from the well-known preference-based utility of von Neumann and Morgenstern. It leads to what we call product utilitarianism. We shall see that product utilitarianism puts some classic discussions of utilitarianism in a different light.

Von Neumann and Morgenstern used preferences between gambles over options, together with the expected utility principle, to measure utilities. This is no part of Bentham's approach. We note that the use of probabilities in this way introduces its own issues, such as those highlighted in the examples of Allais (1953) and of Ellsberg (1961). Whether or not Allais or Ellsberg are thought to present serious difficulties,[2] it may be of interest to present a preference-based representation theorem for utility that does not raise questions of risk because it does not use probability at all. That is what we do here.

We use objective duration together with an integral to measure utility. This is in some ways technically similar to the von Neumann-Morgenstern approach, but there are also notable differences. In particular, there is a natural zero, so that our utility is measured on a stronger scale, a ratio scale, than von Neumann-Morgenstern utility, which is measured on an interval scale. This has consequences deserving of philosophical attention. On our approach, aggregating group utility as a product rather than a sum, and average utility as a geometric average rather than an arithmetic average, is meaningful in the measurement-theoretic sense, although the sum and arithmetic average are not. This is what we call product utilitarianism. On the von Neumann-Morgenstern approach, neither mode of aggregation is meaningful. With regard to some classic issues, we will see (*i*) that product utilitarianism has what are called "prioritarian" consequences, and (*ii*) that the "mere addition paradox" is meaningless, and (*iii*) the "repugnant conclusion" does not follow.

Over his lifetime, Bentham had many views about utility, some contradicting others. The following development focuses on two of his approaches. Part 1 focuses on an abstract extensional account of hedonic measurement. In it a

[1] Another, more prominent, strand is the use of the just noticeable differences of the psychophysics of Weber and Fechner to measure utility. This idea was taken up by Jevons (1871) and by Edgeworth (1879, 1881). This leads to Luce's (1956) theory of semiorders. Our approach here is quite distinct from this idea, which we plan to discuss elsewhere.

[2] For examples of discussion from a large philosophical literature see Buchak (2013) and Seidenfeld (1988).

hedonic episode is characterized by its duration, and by the hedonic intensity, possibly varying, throughout that episode. An episode can just be thought of as a pair consisting of a duration and an intensity function on that duration. Everything else one might think of as relevant is subsumed in the intensity function. Here Bentham's "propinquity" for instance, has no role to play. In our Part II, we allow a richer notion of an episode, where the relative positions of episodes ("propinquity") may be important. This more general approach is useful for modeling decision making and some forms of context effects on utility.

Some related work already exists, but it has largely escaped philosophical attention. We do things differently, in ways closer to Bentham, and develop a more general theory. We discuss related work in Section 6.

1.1 Measurement theory

> *We can tell that one pleasure is greater than another; but that does not help us. To apply the mathematical methods, pleasure must be in some way capable of numerical expression; we must be able to say, for example, that the pleasure of eating a beefsteak is to the pleasure of drinking a glass of beer as five to four. The words convey no particular meaning to us; and Mr. Jevons, instead of helping us, seems to shirk the question. We must remind him that, in order to fit a subject for mathematical inquiry, it is not sufficient to represent some of the quantities concerned by letters.*
>
> Anonymous review of Jevons. In Saturday Review, Nov. 11, 1871 (quoted by Edgeworth, 1887, in *Mathematical Psychics.*)

Utilitarianism sees collective utility as a sum of individual utilities. It was already clear in the 19th century that utilitarianism raises a fundamental question of measurement. But a rigorous and nuanced theory of meaningful measurement was not developed until the 20th century, starting with Scott & Suppes (1958) "Foundational aspects of theories of measurement", and elaborated in Krantz, Luce, Suppes, & Tversky (1971, 1990) *Foundations of Measurement, Vol. I*, Suppes, Krantz, Luce, & Tversky (1990), *Vol. II,* Luce, Suppes, Krantz, Tversky (1990), *Vol. 3*, and Narens (1985) *Abstract Measurement Theory.* Measurement is the mapping of objects onto numbers, in such a way as to reflect some empirical structure. Alternative mappings may reflect the same empirical structure. For example, the empirical structure might simply be an ordering, with the order of the numbers signifying the empirical order. Then a large class of assignments of numbers would equally well represent the same empirical structure. A member of the class can be gotten from another by any order-preserving transformation. A range of different scales of measurement, reflecting more or less empirical content, are possible. For instance, one can think of temperature scales before and after the establishment of an absolute zero. ("Before", it was an *interval scale*—that is, its scale consisted of all transformations of the form $x \to rx + s$, r positive, s real, of any one of its *representations,* for example, the Centigrade representation; "after", it became a *ratio scale*—that is, its scale consisted of all transformations of the form $x \to rx$, r positive, of of any one of its representations, for example, the Kelvin representation.) More empirical structure corresponds to smaller classes of transformations that preserve it. *Meaningful* numerical properties or statistics are those that remain invariant over the appropriate class of transformations. For example, $\frac{^\circ 5}{^\circ 4}$ is meaningful when temperature is measured in degrees Kelvin but not meaningful when measured in degrees Centigrade.

There are scales of measure intermediate in strength between those envisioned in the above epigraph—between mere affixing of labels and a measurement that makes the ratio of 5 to 4 is meaningful. It is evidently of some importance for the discussion of utilitarianism to be clear on what sort of scale utilities are supposed to be measured. But philosophers, who are often so very careful of small details, are not always so careful about this one. For instance, it is sometimes claimed that a utilitarian will prefer larger and larger populations, as the sum of individual utilities is thus increased, even if average quality of life is severely compromised.[3] If, however, utilities are measured on a scale with no meaningful zero, this argument makes no sense. Shifting the scale, so that all utilities are negative, would yield the opposite conclusion. And if zero is not meaningful, then neither is the argument. It is sometimes claimed that utilitarianism would give an unfair advantage to someone who feels with exquisite intensity—a "utility monster".[4] But if individual utilities were measured on scales in which units are conventional, then again, the supposed hypothesis would have no meaning. Evidently, if discussions of utilitarianism are to be done carefully, attention must be paid to the question of measurement. We do so in this article.

Consider a society with a fixed finite number of members, facing a set of alternative social prospects. Each member of society has his or her own utility scale for measuring pleasurable or painful experiences. Traditional utilitarianism—which throughout this article we call *sum utilitarianism*—measures social utility by adding individual utilities. We later propose an alternative to this that we call *product utilitarianism* that eliminates the following meaningfulness difficulty associated with sum utilitarianism.

Suppose both Adam and Eve have either interval utility scales or ratio utility scales, Eve's representation gives prospects A, B, C utilities 3, 2, 1 respectively, and Adam's representation gives them utilities 1, 2, 3. Then the utilitarian sum results in a 3-way tie. But multiplying Adam's utilities by 100 to produce a representation that give A, B, C respectively 100, 200, 300, while keeping Eve's representation the same, would lead to the utilitarian sum reflecting Adam's preferences. And multiplying his utilities by .001 would similarly favor Eve. But the choice of representation is, by hypothesis, arbitrary. In this setting, the "utility monster" is meaningless. So also is the classical utilitarian argument for egalitarianism.

[3]See Parfit (2004). He implicitly assumes a zero point, dividing "lives worth living" from "lives not worth living".

[4]Nozick (1974).

Part I. Modeling Utility Through Hedonic Intensities

2 Utility of Pleasurable Episodes

> *…let us begin with saying: Pleasure is comprised under two dimensions, Intensity and Duration …* (Bentham in Halévy v. 1, Appendix II, p. 302.)

The primary bearers of utility here are *episodes.*[5] They are characterized by the times they begin and end, their intensity and duration of pleasure. Duration is the interval of time between their beginning and end. Intensity need not be constant, it is some function over time. Leaving pain to the side for the moment, the utility of an episode is gotten by "summing up" the constituent pleasure intensities. The utility of an episode is the integral of pleasure intensity with respect to time, evaluated from the beginning to the end of the episode.[6].

We assume that individuals have preferences over episodes, just as von Neumann and Morgenstern assume that individuals have preferences over objective gambles. Where they use chances to measure utilities, we will use objective time.[7] Some of the treatment is the same, but there are also important differences.

If we already have pleasure intensities and time given and preference goes by utility, then some properties of preferences over episodes follow that are deserving of discussion. First, the order of intensities in the episode does not matter:

(1) *Permutability:* If episode 2 comes from episode 1 by permuting two subepisodes of positive duration, then it is a matter of indifference between episode 1 and episode 2.

This may appear to fly in the face of the commonplace that order of experiences makes a difference in judged overall pleasure. The Benthamite can reply that this confuses order of the experiences that engender pleasure with the order of pleasure intensity. You may like appetizer before entree before main before dessert better than permutations of courses, but this just shows that permutations of courses do more than permute subintervals of pleasure intensity: They change the intensities of pleasure within those sub-intervals. We deal with noncommutativity and operations that correspond to changes in pleasure intensities in Part II.

[5] Here we follow the terminology of Kahnemann, Wakker and Sarin (1997) "Back to Bentham."

[6] Bentham does not have the notion of the integral, but the idea is clear in chapter II of Jevons' Theory of Political Economy, in Edgeworth's (1879) Mind article and in Edgeworth's appendix III to Mathematical Psychics, quoted later in Section 6

[7] A different strand of Bentham's thought uses subjective "atoms" of time. That idea leads in a different direction. Edgeworth toys with the idea of using both subjective and objective time.

(2) *Ordering: Preference orders all episodes.*

(3) *Independence:* Suppose episode 2 is preferred to episode 1, episode 1 is a sub-episode of episode 3, and episode 4 comes from episode 3 by substituting episode 2 for episode 1. Then episode 4 is preferred to episode 3. [Similarly for indifference.] If an episode is of constant intensity, for a partition of it into n sub-episodes of equal duration there is indifference between sub-episodes, for any n. Once again, this Independence assumption is valid because substitution consists of the substitution of experiences and not episodes. This exemplifies the extensional character of the modeling of this part of the article.

(4) *Average:* For any episode, there is a constant episode of the same duration with the same utility. Between them, it is a matter of indifference.

(5) *Null Episode:* Every episode of positive intensity and positive duration is preferred to the null episode of zero duration.

2.1 Measurement of pleasure

The current situation is a special case of a more general theory of utility presented in Part II. For the current situation, we can measure pleasure of episodes as follows. Pick a positive intensity and a positive time period. An episode of that duration at that intensity will serve as a unit of pleasure. The same intensity for two time units is two units of pleasure; the same intensity for half the time is half a unit of pleasure. These special episodes are on a ratio scale with a distinguished zero (any instantaneous interval of pleasure) and an arbitrary unit, like the meter unit in the measurement of length. Other episodes are put on the scale, by matching them with one of these where, as is experienced by the individual, it is a matter of indifference.

We measure intensity of pleasure as follows. Start with a base unit episode as above. If an episode of half the time at constant intensity is of equal utility as the base unit episode, then that intensity is twice the intensity of the base unit; if an episode of twice the time at constant intensity is of equal utility as the base unit, then that intensity is half of that of the base unit. In this way, the intensity of pleasure in any pleasure episode can be measured. As an result, the intensities of pleasure form a ratio scale, because any pleasure episode can be taken as the unit.

The above measurement of pleasure relies on the the existence of constant pleasure episodes for each constant intensity and for each duration. The theory of pleasure presented Part II does not make these assumptions.

3 Pain

The second of Bentham's "sovereign masters" is pain. In the continuation of the passage previously quoted, he uses preferences to compare pains. Which would a subject rather avoid?

> If of two pains a man would as lief escape one as the other, such two pains must be reputed equal. . . . (Bentham in Halévy v. 1, Appendix II, p. 302.)

The procedure outlined above can be just as well applied to give a ratio scale for pains. Since pleasurable episodes are preferred to painful ones, purely pleasurable and purely painful episodes can be measured on a common scale, with pleasures having positive numbers and pains having negative ones.

Note however, that now we have made two arbitrary choices of units, one for pleasure and one for pain. We could choose the same time interval for each choice of unit, since time is objective. But we have no way of starting out by choosing the same intensity—we have no way, at this point of saying that a given intensity of pleasure is of the same magnitude as a given intensity of pain. Our measurement of intensities presupposes our choice of unit. Thus the single scale for pleasure and pain that we have so far is not quite a ratio scale: it depends on two arbitrary choices of units, not one.

4 Pleasure and Pain Together

We may have episodes that combine both pleasure and pain. How are they to be treated? First we have to ask what kinds of combinations are possible. We may have episodes that are pleasurable for a stretch of time and painful for a stretch. Can we have also episodes that are both pleasurable to some extent and painful to some extent at the same time? This seems a clear psychological possibility. We can proceed either way.

How do pleasure and pain interact in determining the utility of an episode? Is it possible that someone might prefer a pleasure with a small amount of pain to the pure pleasure? Bentham would say "No," and insist that the interaction of pleasure and pain is purely additive. Thus,

> If of two sensations, a pain and a pleasure, a man would as lief enjoy the pleasure and suffer the pain, as not enjoy the first and not suffer the latter, such pleasure and pain must be reputed equal . . . (Bentham in Halévy v. 1, Appendix II, p. 302)

and

> Sum up all the values of all the pleasures on the one side, and those of all the pains on the other. The balance, if it be on the side of pleasure, will give the good tendency of the act upon the whole, with respect to the interests of that individual person; if on the side of pain, the bad tendency of it upon the whole.[8]

[8] Bentham, *An Introduction to the Principles of Morals and Legislation Chapter IV: "Value of a Lot of Pleasure or Pain, How to be Measured",* section V. The setting here is different. Bentham is thinking here of just noticeable differences. But the principle is clear.

If so, we can use mixed episodes to align the units of the pleasure and pain scales, as Bentham suggests in the first passage. Assume we have chosen units for each. If an episode consisting pleasure with pleasure intensity 1 for two units of time together with pain with pain intensity 1 for one unit of time is a zero on both scales, we can say that one of our pain units is equal to two of our pleasure units. We now have common ratio scale for episodes of pleasure, pain, and mixtures of the two.

This is all under the assumptions of additivity of pleasures and pain above made by Bentham. In Part II we show how to derive these additivity assumptions from qualitative properties of episode measurement. One consequence of additivity is, as Bentham says, that a pleasure may be able to be added to a pain to produce to produce an event that is neither pleasurable or painful—a neutral event— that has utility 0.

Suppose A is a pleasurable event and there are painful events B and D can follow A such that A immediately followed by B is pleasurable and A immediately followed by D is painful. Then, because of the assumed continuity and the wide variety of painful events, it is reasonable to postulate the existence of an painful event C between B in D in terms of painfulness such that A immediately followed by C is neutral. Such C must have utility 0.

This is how Bentham and other utilitarians thought about 0 utility. It is a valid concept for an individual. But, one may ask whether the identification of 0 utilities across individuals is a valid comparison. Why should one person's neutral events defined in this manner be identified with another's neutral events?

But it is also true on utilitarian principles that the null episode, the episode of zero duration, has utility zero, as we note above in (5). The *intensities* of Adam and Eve's instantaneous experiences may disagree. But episodes of shorter and shorter durations at those intensities must have utilities that converge to 0. Zero utility has the same meaning for each individual.

5 Aggregation of Pleasures

Bentham thought that the utility for a group should be measured as the sum of the utilities of its members. Consider a stretch of time, where the members of the group remain constant. How can utility of the group be meaningfully quantified? On the foregoing account of utility—as on the von Neumann–Morgenstern account—it cannot be as a sum, because individual utilities are only measured up to an arbitrary unit. Multiply Peter's units one constant and Paul's by another, and provided their interests conflict, you may reverse the pair's group preferences. But the foregoing account, unlike von Neumann's-Morgenstern's, has a distinguished zero, and measures each individual's utilities on a ratio scale. This is similar to the situation encountered by psychophysicists. However, their situation is different in that it concerns measurements of only positive sensations, which here translates into measurements of episodes composed of only pleasure intensities. (It analogously applies to measurements of episodes composed of only painful intensities. It does not apply to mixes of such episodes or mixes including zero episodes.) To avoid the difficulty with the nonmeaningfulness of sums across individuals, Stevens (1948) suggested using the geometric mean of measurements. This has the property that it preserves the numerical ordering of the aggregates of individual episodes no matter which representation from individuals' ratio scales are used. It is also a ratio scale in the sense that if each individual representation is multiplied by the same positive constant then the aggregate representation is multiplied by that constant. Measurement theorists (e.g., Aczél & Roberts, 1989, Corollary 3.1) have shown the following: *The only scales that preserve the numerical ordering of the aggregates no matter which representation from ratio scales are used by individuals are strictly monotonic transformations of the geometric mean.* An example of this is a scale that is formed by suggestion at the beginning of this section that is obtained by multiplying together representations from each individual's scale.

Does the use of the product (or geometric mean) for aggregation make much of a difference from the hypothesized utilitarian sum? Here we briefly note some consequences that may be of philosophical interest. Many influential criticisms of traditional utilitarianism assume with Bentham that utilities can be measured on a cardinal scale and then added up. According to one criticism, the utilitarian sum leads to the conclusion that a social state that gave Adam 100 utiles and Eve zero, would be as good as one that gave each 50 utiles. "Prioritarians", Parfit (1991), think that this is wrong. Another criticism, in Parfit (1984), says that for utilitarians a social state with huge numbers of people living lives each with positive utility near zero is better than one with a modest number of people living well. For product utilitarians, neither of these conclusions follows. We return to these considerations when we revisit aggregation in Part 2 of this paper.

6 Related Work for Part 1

The modeling of utility through pleasure intensities described above can be considered a formulation of Edgeworth's hedonimeter with modern ideas from measurement theory and psychophysics. Edgeworth (1881) writes:

> To precise the ideas, let there be granted to the science of pleasure what is granted to the science of energy ; to imagine an ideally perfect instrument, a psychophysical machine, continually registering the height of pleasure experienced by an individual, exactly according to the verdict of consciousness, or rather diverging therefrom according to a law of errors. From moment to moment the hedonimeter varies; the delicate index now flickering with the flutter of the passions, now steadied by intellectual activity, low sunk whole hours in the neighbourhood of zero, or momentarily springing up towards infinity. The continually indicated height is registered by photographic or other frictionless apparatus upon a uniformly moving vertical plane. Then the quantity of happiness between two epochs is represented by the area contained between the zero-line,

> perpendiculars thereto at the points corresponding to the epochs, and the curve traced by the index ; or, if the correction suggested in the last paragraph be admitted, another dimension will be required for the representation. The integration must be extended from the present to the infinitely future time to constitute the end of pure egoism.

Sarin & Wakker (1997) formulate a generalization of hedonimeter for use in decision theory. Their primitive concepts are episodes with ordinal instantaneous hedonic pleasures, pains, and neutral (zero intensity) experiences, $f(t)$, that varies with time t throughout the episode's duration, and a weakly ordered preference relation $\precsim$ over episodes that represents the obvious ordering of the amount of pain or pleasure in the episode. For example, when A and B are both pleasurable episodes, $A \precsim B$ is read as, "The amount of pleasure produced by A is less than or the same as the amount of pleasure produced by B." Sarin & Wakker provide axioms in terms of these primitives and show that their axiomatization is logically equivalent to the following, where F and G are arbitrary episodes spanning time intervals $[a, b)$ for F and $[c, d)$ for G and with $f(t)$ being the instantaneous hedonic function for F and $g(t)$ for G,

$$F \precsim G \text{ if and only if } \int_a^b e^{-rt} v[f(t)]dt \precsim \int_c^d e^{-st} w[g(t)]\,,$$

where v and w are unique up to multiplication by a positive constant and r and s are uniquely determined if v is nonconstant. They interpret e^{-rt} and e^{-st} as nonconstant time discounting factors.

In our modeling below of preference for hedonic episodes, we employ different axioms. In particular our axiomatization does not use instantaneous hedonic functions like $f(t)$ above.

Part II. Modeling Utility Through Preference for Hedonic Episodes

7 Individual Preference Modeling

> *Considered with reference to an individual, in every element of human happiness, in every element of its opposite unhappiness, the* elements*, or say dimensions of value (it has been seen,) are four: intensity, duration, propinquity, certainty; add, if in a political community, extent. Of these five, the first, it is true, is not susceptible of precise expression: it not being susceptible of measurement. But the four others are.*
>
> From Bentham's (1822) "Codification Proposal", p. 11.

7.1 Introduction

For Bentham, the dimension of certainty is easy to handle: The utility of an episode E happening with probability p is just the utility of E happening multiplied by the probability p. As mentioned at the beginning of this article, Bentham did not have beyond this a risk component to his utility theory, and, in this part of the article, probabilistic concerns will be ignored. The dimension of extent, is concerned with "the number of persons to whom [happiness] *extends*".

Bentham modeled propinquity as a factor that multiplies by a positive number the utility of an immediately experienced version of the episode. Thus for future episodes it is like a discount factor, except it can also lead to an increase of utility. But the Benthamites did not deal with the subtleties of how a future version of an episode is related to an immediately experienced one. Will it be experienced in the same way in the future? Perhaps not, because it has a different past leading up to it. It is this context effect of dependence on the past that produces problems for hedonic decision making.

This section considers an episode's dependence on the past to be another dimension of value. In particular, an episode is considered to be a physical entity that has a beginning, and associated with this beginning is a context representing the individual's past hedonic experiences. Two episodes are then defined to be *hedonically similar* if and only if they are physically identical except for physical features that are irrelevant for hedonic calculation, have the same overall amounts of pleasure (or pain) associated with them, but have different pasts. Hedonism will be measured in terms of possible hedonic episodes that the individual could have experienced. The existence of hedonically similar ones increases the range of possible experiences, and gives decision making the flexibility that people find useful and employ. For example, it allows for comparisons of judgments of the amounts of hedonism in issues like, "Is eating the main course first and dessert second is more pleasurable than eating the dessert first and the main course second?".

7.2 Hedonic episodes

Throughout this section, $\mathcal{H}$ will denote the set of hedonic episodes. Each hedonic episode H spans a finite interval of physical time, $[a, b)$, and beginning time a, and b being the beginning time of the *next* episode that immediately follows H. We also consider as episodes, the instantaneous episode [a], viewed as a limit of episodes $[a, x)$ as the time of $x \to$ the time of a. $[a]$ will be assigned the hedonic value 0.

We assume a preference ordering $\precsim$ on the *amount of hedonism* produced by each episode in $\mathcal{H}$. For hedonic events G and H with H being instantaneous, G is said to be *pleasurable* if and only if $H \prec G$, G is said to be *neutral* if $G \sim H$, and G is said to be *painful* if and only if $G \prec H$. Thus for pleasurable G, neutral H, and painful K, it follows from $\precsim$ being a weak order that $K \prec H \prec G$.

Bentham wanted to develop for law, economics, and political theory a mathematical foundation and theory along the lines of the physics of his time. An adequate theory of physical measurement didn't exist then. It was later developed by Helmholtz in 1887 and improved upon by Hölder in 1901. Bentham developed his own approach for measuring hedonism, which has been formalized in Part I of this article as an integral of hedonic intensities. As quoted in the epigraph at the beginning of this section, Bentham realized that there was, from his perspective, a measurement issue with this approach: "Of these five, the first [intensity] is not susceptible of precise expression: it not being susceptible of measurement." This becomes a non-issue when we apply the Helmholtz-Hölder approach to the measurement of hedonism.

The hedonic episode H is said to be a *physical concatenation* of the hedonic episodes F and G, in symbols $H = F^\frown G$, if and only if

(*i*) F and G are sub-episodes of H,

(*ii*) F and G have no durations in common, and

(*iii*) the union of the durations of F and G is the duration of H.

Note that if F, G, and K are hedonic episodes and either $(F^\frown G)^\frown K$ or $F^\frown(G^\frown K)$ are defined, then

$$(F^\frown G)^\frown K \sim F^\frown(G^\frown K)\,. \tag{1}$$

7.3 Additive representations

The Helmholtz-Hölder theory of the measurement of physical dimensions is that each fundamental physical dimension like distance, time, mass, charge, etc., had a physical concatenation operation O defined on it. For example, length was measured in terms of rigid measuring rods, and rods e and f could be abutted together to form a new rod $e\,\mathrm{O} f$ whose length would be the sum of the lengths of e and f. We do the same for the amounts of hedonism produced by episodes A and B.

Define the multivalued operation of *formal concatenation,* $\oplus$, on the set of episodes $\mathcal{H}$ as follows: For all A, B, and D in $\mathcal{H}$,

$$A \oplus B \sim D$$

if and only if there exist episodes F, G, and H such that

$$A \sim F,\ B \sim G,\ D \sim H \text{ and } F^\frown G = H\,.$$

Consider the episode $(A \oplus B) \oplus D$, where A, B, and D are arbitrary elements of $\mathcal{H}$. By the definition of $\oplus$, let J, K, and L in $\mathcal{F}$ be such that

$$J \sim A,\ K \sim B, \text{ and } L \sim D$$

and

$$A \oplus B \sim J^\frown K \text{ and } (A \oplus B) \oplus D \sim (J^\frown K)^\frown L\,.$$

It then follows from Equation 1 that $J^\frown(K^\frown L)$, and thus that

$$\textit{Associativity of } \oplus\text{:} \quad (A \oplus B) \oplus D \sim A \oplus (B \oplus D)\,. \tag{2}$$

Helmholtz and Hölder axiomatized fundamental physical dimensions in terms of qualitative properties of their concatenation operations and their qualitative orderings. The orderings compared sizes, e.g., by laying measuring rods side by side and see which is longer by seeing which spanned the other. We follow them for measuring amounts of hedonism by providing axioms in terms of the qualitative ordering $\precsim$ and operation $\oplus$. Even though the dimension of hedonism is psychological and not physical, its measurement will obey the same measurement principles as physical measurement, achieving an ideal goal of Bentham.

The qualitative structure $\langle \mathcal{H}, \precsim, \oplus \rangle$ is assumed to satisfy the following six axioms for all A, B, and D and E in $\mathcal{H}$:

1. $\oplus$ *is a weak operation:* There exists an episode F such that $A \oplus B \sim F$.

2. *Order Density:* If $A \prec B$ then for some episode F, $A \prec F \prec B$.

3. *Existence of Negative Elements:* There exist an episode $-A$ and a neutral element Z such that $A \oplus -A \sim -A \oplus A \sim Z$.

4. *Neutrality of* $\oplus$*:* If A is neutral, then $A \oplus B \sim B \oplus A \sim B$.

5. *Monotonicity of* $\oplus$*:*

$$A \precsim B \text{ iff } A \oplus D \precsim B \oplus D \text{ iff } \mathcal{D} \oplus A \precsim D \oplus B\,.$$

6. *Dedekind completeness*: Each $\precsim$ bounded nonempty subset of $\mathcal{H}$ has a $\precsim$ least-upper bound.

Definition. A function φ into the positive reals is said to be an *additive representation* for $\mathcal{H}$ if and only if the following two statements hold for all A, B, and D in $\mathcal{H}$:

1. $A \precsim B$ iff $\varphi(A) \leq \varphi(B)\,.$

2. $\varphi(A \oplus D) = \varphi(A) + \varphi(D)\,.$

The weak ordering of $\precsim$, the associativity of $\oplus$ (Equation 2), and the above six assumptions say that the qualitative structure $\mathfrak{H} = \langle \mathcal{H}, \precsim, \oplus \rangle$ is a Dedekind complete weakly ordered group that is order dense. A famous theorem of mathematics by Hölder shows that such groups have additive representations.

Theorem 1. *The set of additive representations for* $\mathcal{H}$ *is a ratio scale, that is,*

(1) there exists an additive representation for $\mathcal{H}$*; and*

(2) for all additive representations φ and ψ of $\mathcal{H}$, there exists a positive real s such that $\psi = s\varphi$.

Theorem 1 requires $\oplus$ to be a weak operation, which combined with its other axioms, require the additive representations to be onto the real numbers. There are more general versions of Hölder's Theorem that also apply when $\mathfrak{H}$ is bounded. In such bounded cases, $\mathfrak{H}$ has a ratio scale of additive representations such that each representation is onto a bounded interval of reals.

Let $\mathcal{S}$ be the ratio scale of additive representations for $\mathcal{H}$, A be in $\mathcal{H}$, and φ a representation in $\mathcal{S}$. $\varphi(A)$ is called the *utility* of A, and the quantity $\alpha(A)$ satisfying the equation,

$$\varphi(A) = \alpha(A) \cdot (\text{the measurement of the duration of } A)\,,$$

is called the *average utility of A.* This looks similar to the utility of an outcome c of a gamble that has probability p occurring, i.e.,

$$u(c, p) = u(c) \cdot p\,,$$

with $u(c)$ taking the place of average utility and p taking the place of duration.

Part 1 showed that Bentham's measurement of utility consisted of taking the utility of an episode as the integral of pleasure-pain intensity with respect to time, evaluated from the beginning to the end of the episode. For Bentham's kind of hedonic episodes F and G, define $\precsim$ as follows with $b(X)$ being "the time that is the beginning of X" and analogously for the end of X, $e(X)$:

$$F \precsim G \ \text{ if and only if } \int_{b(F)}^{e(F)} F \ \leq \ \int_{b(G)}^{e(G)} G\,.$$

Then Bentham's measurement of pleasure is a special case of Theorem 1.

Two concatenation operations have been used for combining amounts of hedonism, the physical concatenation operation, $\frown$, and the formal concatenation operation, $\oplus$. $\frown$ is non-commutative, that is, in general,

$$A^\frown B \not\sim B^\frown A\,.$$

because in "$B^\frown A$", A's immediate past is B, whereas in "$A^\frown B$ it is generally something else. In other words, if $A^\frown B$ exists, then $B^\frown A$ cannot. Instead, an appropriately hedonically similar episode A' has to be chosen so that $B^\frown A'$ is well-formed. But, in general, A', when calculating the the amount of hedonism in $B^\frown A'$, will have a different amount of hedonism than A does in $A^\frown B$, because A' has B in its immediate past while A does not. This shows

$$A^\frown B \not\sim B^\frown A'\,.$$

However, it follows from Theorem 1 that

$$A \oplus B \sim B \oplus A\,.$$

This difference in commutativity reflects that past context is preserved in the calculation of the physical concatenations of amounts of hedonism, but in formal concatenation, while the past contexts are used in the calculations of the amounts of hedonism from A and B, the context is lost in the calculation of their formal concatenation, $A \oplus B$. This allows Benthamite utility based on $\precsim$ and $\oplus$ to share the additivity properties of utility from current economics.

8 Aggregation Continued

Now that we have a general theory of utility of episodes, we can revisit the question of aggregation in a more general setting. Suppose that a group is contemplating alternative courses of action that will affect the utilities of its members. And suppose that for each course of action under consideration the net utility of each of the members is positive. Then the product utility, or alternatively the geometric mean, is meaningful. Utilities of members are on a ratio scale, and rescaling intervals does not change the group ranking. And it retains the prioritarian flavor on which we remarked at the beginning of this article. On the other hand, perhaps a disaster is at hand and courses of action under consideration all give each member an excess of pain over pleasure. Here the previous case is inverted. Each individual is on a ratio scale of pain. Aggregation by the product is still meaningful in exactly the same way. But priotarianism is lost, and indeed reversed. What about the mixed case, where in some courses of action some individuals enjoy net pleasure, while others suffer net pain? We have to say here that there does not seem to be a plausible utilitarian avenue to follow. Cancelling pains against pleasures at the aggregate level is simply not meaningful. If we independently rescale the units for different individuals, we can change the social ordering of alternatives. But the units don't mean anything in our scheme of measurement. Trying an alternative aggregation scheme that is meaningful, such as product of individuals enjoying pleasures divided by product of individuals suffering pains, gives morally crazy results. We see product utilitarianism as a approach that is only viable in a restricted domain. This should not disqualify it from consideration.

It is a recurrent theme in social choice theory that some desirable property of social preference must be sacrificed to achieve a consistent aggregation of individual preferences. Here we give up what Arrow (1951) calls "unrestricted domain" by restricting the individual preferences to be over only pleasurable events or over only painful ones, but not mixes, where for some event one individual finds it pleasurable and another finds it painful. Utilitarian sum aggregation across all events would require interpersonal comparisons of utility that, in their more skeptical moments, the Utilitarian founding fathers themselves found dubious. Thus, in a now famous passage in an unpublished manuscript found by Halévy, Bentham writes:

> Tis vain to talk of adding quantities which after the addition will continue distinct as they were before, one man's happiness will never be another man's happiness; a gain to one man is no gain to another;

> one might as well pretend to add 20 apples to 20 pears, after which you had done that could not be 40 of any one thing but 20 of each just as they were before. This addibility of the happiness of different subjects, however, when considered rigorously it may appear fictitious, is a postulatum without the allowance of which all political reasoning is at a stand. (From Halévy v3 *p. 349.*)

And, Jevons (1881) writes,

> The reader will find, again, that is never, in any single instance, an attempt made to compare the amount of feeling in one mind with that in another. I see no means by which such comparison can be accomplished. The susceptibility of one mind may, for what we know, be a thousand times greater than that of another. But, provided that the susceptibility was different in a like ratio in all directions, we should never be able to discover the difference. Every mind is thus inscrutable to every other mind, and no common denominator of feeling seems to be possible. (*p. 21*)

Without interpersonal comparability of units, we see that, nevertheless, in some important classes of cases meaningful aggregation is possible. Even with the restriction to only positive or only negative utilities, the product or geometric mean aggregation principle can have important application. A Prioritarian flavor is evident in the following examples.

Suppose that a windfall has been found and the feasible social options under consideration all give each member of the group positive utility. Then we can use the product to aggregate. For instance, new trees appear in the garden of Eden, and there is new fruit to distribute. Distribution (A) gives Adam utility 1 on one version of his ratio scale, and Eve 20 on one version of hers, while distribution (B) gives Adam 5 and Eve 5. We resist the urge to look at the sum which, as Jevons says, is meaningless; we look at the product. Then (B), with a product of 25 is socially preferable to (A) with a product of 10. If we multiply Adam's utilities by one positive constant and Eve's by another, (B) is still preferable to (A). Note that by choosing the constants, we could make (A) look more egalitarian than (B) because "egalitarian" doesn't mean anything in this framework. Suppose that we multiply Adam's utiles by 20, and leave Eve's alone. Then, in this representation, (A) looks egalitarian, but Adam does so well in (B) that the aggregate good favors (B). In this representation, (B) has an aggregate utility of 500, while (A) has one of 400.

If we know that Adam's utilities (on some version of his ratio scale) is a function of the quantity of some real or monetary good possessed, and likewise for Eve, then we can do more. Consider the case of dividing \$100 between Adam and Eve, with the proviso that each must get at least \$1. On some choice of units for their ratio scales, Adam's utility function is $\varphi_a(\$x) = x$ and Eve's is $\varphi_e(\$x) = \sqrt{x}$. In this case, if the utilitarian sum were meaningful, the only utilitarian sum solution would be \$99 to Adam and \$1 to Eve. The utilitarian product solution is $\frac{2}{3}\$100$ to Adam and $\frac{1}{3}\$100$ to Eve.[9] Adam receives more money than Eve because Eve has a faster diminishing marginal utility for money than Adam.

8.1 Parfit's counterexamples

The foregoing examples all deal with a fixed population. If alternative scenarios being evaluated involved different populations, Parfit raised a difficulty for Utilitarianism thus:

> For any possible population of at least ten billion people, all with a very high quality of life, there must be some much larger imaginable population whose existence, if other things are equal, would be better even though its members have lives that are barely worth living. Parfit(1984) p. 388

Parfit was addressing a Utilitarian Sum, assuming that utilities are real numbers, that "lives barely worth living" identifies some positive real number, and using the Archimedean property of the reals. He has another argument, in the same framework, against those who would compare populations using the arithmetic average. A population with a few extremely happy people has an average utility higher than one which, in addition, has many people who are almost, but not quite, as happy.

> Suppose that Adam and Eve lived these wonderful lives. On the Average Principle is would be worse if, *not instead but in addition*, the billion billion other people lived. [Note: Specified earlier as having a quality of life almost as high.] This would be worse because it would lower the average quality of life. Parfit(1984) p. 420.

If utility is measured on an absolute scale, then given his assumptions, Parfit's examples are correct and telling.

On our account, individual utilities are measured not on an absolute scale but on a ratio scale. As we have seen, for a fixed population the utilitarian sum and arithmetic average are not meaningful, but the product and geometric mean are. What about comparisons of utility across populations? Do measurement considerations still support Parfit's arguments?

First consider the product. With product utilitarianism, adding lives (considered as episodes) whose utility is greater than one increases aggregate utility, while adding lives whose utility is between between zero and one, *decreases* the

[9] This is gotten by finding the maximum of the utilitarian product $(100-x)\cdot\sqrt{x}$ by setting its derivative $= 0$, that is,

$$0 = \frac{d}{dx}[(100-x)\cdot\sqrt{x}] \;=\; -\sqrt{x} + (100-x)\frac{1}{2}\frac{1}{\sqrt{x}} \;=\; -2x + (100-x)\,,$$

and thus $x = \frac{1}{3}$.) Then by letting, $S =$ "\$99 to Adam and \$1 to Eve" and $P =$ "$\frac{2}{3}\$100$ to Adam and $\frac{1}{3}\$100$ to Eve", we see that the utilitarian sum is indifferent between S and P while the utilitarian product prefers P to S.

aggregate utility. Adding a life with utility 2 doubles the product; adding a life with utility .5 cuts it in half. But utility of 1 or any positive real is not meaningful when utilities are measured on a ratio scale, so it is not meaningful to ask whether adding a life with positive utility increases of decreases the aggregate.

Now consider the geometric mean. On one representation, Adam has utility 101. We could add Eve who, on one representation, would have utility 100. This would decrease the geometric mean so, by Parfit's second argument, it would argue for leaving Adam alone. But Eve's utilities could just as well be rescaled to 1000, which would increase the geometric mean. Or to 101, which would leave it unchanged. Likewise for all those other people. In our measurement setting, both of Parfit's arguments fail to be meaningful. It is not the size of a particular product or geometric mean that is important, but their comparison for the *same* population.

In our treatment, zero utility is defined in terms of the null episode or equivalently, and instantaneous one. It is a further assumption that this divides lives barely worth living from those not. One could contemplate a different value for "lives barely worth living". This does not help, but rather introduces additional problems. Perhaps lives of pain might be worth living for some, and lives of pleasure not worth living for others. In such a mixed population, with these considerations in play, we have no meaningful mode of aggregation at all.

In all the foregoing cases the utility comparisons used in the examples are meaningless.Without further assumptions the theory is silent. The type of utility measurement developed here preserves some of the conclusions of naive classical utilitarianism, but not all of them.

9 Discussion

We have developed a theory of utility measurement along lines suggested by some of Bentham's writings. This is based on preferences on episodes, rather than intensities. There are no objective lotteries as in von Neumann-Morgenstern. Instead there are objective durations. Our approach gives us a natural zero. Thus utilities are measured on a ratio scale, which is stronger than the interval scale gotten by von Neumann-Morgenstern, but weaker that the absolute scale sometimes envisioned by Bentham, and by some of his critics.

We develop our theory in two stages. In each, individuals have pleasure and pain measured on a ratio scale. The first stage formulates ideas outlined by Bentham and later more fully developed by Edgeworth through his concept of a "hedonometer". It describes the amount of pleasure accumulated in an episode. However, this is not flexible enough for most issues involving individual decision making, particularly those that rely on the comparison of hypothetical or contextual episodes. Using a different form of measurement, based on preference for amounts of happiness and methods from physical measurement, the second stage formally captures the first as a special case and is able to deal with hypothetical and some forms of contextual episodes.

The way utility is measured has consequences for meaningful aggregation of utilities, and thus for utilitarianism. The concept of “meaningful aggregation” is often not considered and meaningless aggregation procedures abound throughout the Utilitarianism literature. In the case where ratio scaled utilities of the members of a group are always positive or always negative, aggregation by product is meaningful, although aggregation by sum is not. This has consequences for philosophical discussions of Utilitarianism. A number of famous thought experiments are simply meaningless. And in the positive realm Product Utilitarianism has Prioritarian consequences that seem to have escaped discussion.

References

Aczél, J., & Roberts, F. S. (1989). On the possible merging functions. *Mathematical Social Science, 17,* 205–243.

Allais, M. (1953).Le comportement de l’ homme rationnel devant le risque: critique des postulats et axiomes de l’ école Américaine. *Econometrica, 21* 503–546.

Arrow, K. (1951). *Social Choice and Individual Values.* New York: Wiley.

Bentham, J. (1789). *An introduction to the Principles of Morals and Legislation.* London:

Bentham, J. (1822). Proposal, Codification. “Addressed by Jeremy Bentham to All Nations Professing Liberal Opinions, 1822.” *Works, IV,* 535–94.

Bentham, J. *The Works of Jeremy Bentham.* Published under the Superintendence of his Executor, John Bowring (Edinburgh: William Tait, 1838-1843). 11 vols. Vol. 4. 8/20/2017. http://oll.libertyfund.org/titles/1925#Bentham_0872-04_4619.

Buchak, L. (2013) *Risk and Rationality.* New York and London:Oxford University Press.

Edgeworth, F. Y. (1879). The Hedonical Calculus *Mind, 4* 394–408.

Edgeworth, F. Y. (1881). *Mathematical Psychics.* London: G. Kegan Paul, Appendix III “On Hedonimetry”.

Halévy, E. (1901, 1995). *La formation du radicalisme philosoplique I,II,III. (nouvelle édition réviséé, 1995).* Paris: Presses Universitaires de France.

Helmholtz, H. v. (1887). Zählen and Messen erkenntnistheoretisch betrachtet. *Philosophische Aufsätze Eduard Zeller gewidmet.* Leipzig.

Jevons, W. S. (1871). *Theory of Political Economy.* Oxford: Oxford University.

Kahnemann, D., Wakker, P., & Sarin, R. (1997). Back to Bentham: Explorations of Experienced Utility. *Quarterly Journal of Economics, 112,* 375–405.

Krantz, D. H., Luce, R. D., Suppes, P. & Tversky, A. (1971). *Foundations of Measurement, v.1.* New York: Academic Press.

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

Luce, R. D., D. H. Krantz, P. Suppes, & A. Tversky (1990). *Foundations of Measurement, Vol. III.* New York: Academic Press.

Narens, L. (1985). *Abstract Measurement Theory.* Cambridge, Mass.: The MIT Press.

Nozick, R. (1974) *Anarchy, State and Utopia.* New York: Basic Books.

Parfit (1984). *Reasons and Persons.* Oxford: Clarendon Press.

Parfit (1991). *Equality or Priority.* The Lindley Lecture: University of Kansas.

Parfit, D. (2004) Overpopulation and the Quality of Life. In J. Ryberg and T. Tännsjö (Eds.), *The Repugnant Conclusion,* 7–22.

Sarin, R., & Wakker, P. (1997). Benthamite utility for decision making. Working paper, CenTER, Tilburg, The Netherlands.

Seidenfeld, T. (1988). Decision Theory without 'Independence' or without 'Ordering'. *Econonics and Philosophy, 4,* 267-290.

Scott, D. and Suppes, P. (1958). Foundational aspects of theories of measurement. *Journal of Symbolic Logic, 23,* 113–128.

Suppes, P., D. H., Krantz, R. D. Luce, & A. Tversky (1990). *Foundations of measurement, Vol. II.* New York: Academic Press.

von Neumann, J., & Morgenstern, O. (1947) *Theory of Games and Economic Behavior, 2nd ed,* Princeton:Princeton University Press.