MBS TECHNICAL REPORT 18-04

Utilitarianism from the Perspective of Modern Psychology*

Louis Narens[†] Brian Skyrms[‡]

Abstract

Utilitarianism is a philosophical theory that flourished from the late 18th to early 20th centuries and dominated in England economic, social, and political thought. It was based on psychological principles and led the way to current neo-classical economic theory. Its psychology presented foundational problems in the scientific measurement of individual and group utilities. This together with the infusion of positivist ideas and methods into the early 20th century economic, social, and behavioral sciences led to its demise. Starting in the 1950s, new techniques to measurement and scientific laws, especially qualitative axiomatizations, invariance principles, and "scientific meaningfulness" methods, were developed by psychologists; for example, those summarized in *Foundations of Measurement, Vol. I* (1971) by Krantz, Luce, Suppes, and Tversky, *Abstract Measurement Theory* (1985) and *Theories of Meaningfulness* (2002) by Narens, and *On Meaningful Scientific Laws* (2015) by Falmagne and Doble. These provided among other things new and rigorous approaches to the measurement of subjective intensity and thus to the measurement of utility in particular. This article explores many of the utility concerns that contributed to Utilitarianism's downfall and argues that they are greatly alleviated by achievements in modern measurement theory.

1 Introduction: The Rise and Fall of Utilitarianism

The Rise

Nature has placed mankind under the governance of two sovereign masters, pain and pleasure. It is for them alone to point out what we ought to do, as well as to determine what we shall do. On the one hand the standard of right and wrong, on the other the chain of causes and effects, are fastened to their throne. They

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

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

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

govern us in all we do, in all we say, in all we think: every effort we can make to throw off our subjection, will serve but to demonstrate and confirm it.

Jeremy Bentham, Ch. 1, *Introduction to the Principles of Morals and Legislation,* 1879.

The Fall

If the psychological critics of Economics had troubled to do these things they would speedily have perceived that the hedonistic trimmings of the works of Jevons and his followers were incidental to the main structure of a theory which—as the parallel development in Vienna showed—is capable of being set out and defended in absolutely non-hedonistic terms. As we have seen already, all that is assumed in the idea of the scales of valuation is that different goods have different uses and that these different uses have different significances for action, such that in a given situation one use will be preferred before another and one good before another. Why the human animal attaches particular values in this behaviouristic sense to particular things, is a question which we do not discuss. That may be quite properly a question for psychologists or perhaps even physiologists. All that we need to assume is the obvious fact that different possibilities offer different stimuli to behaviour, and that these stimuli can be arranged in order of their intensity.

Lionel Robbins, *An Essay on the Nature and Significance of Economic Science,* 1932, p. 86.

From the late 18th to the early 20th centuries, Utilitarianism played a major role in the shaping of economic theory, in ethics, and in a movement for social and reform to enhance the utility of the populace. It formulated its theory in terms of the simple psychology of seeking pleasure and avoiding pain. This psychology gave rise to a utility function that measured an individual's pleasure and pain. In terms of properties of utility functions, the Utilitarians sought to described the consequences of individuals economic behavior and to provide the foundations for establishing a moral society based on scientific principles. The goal was to maximize the happiness of a population of people, given available resources. On this basis they argued for universal suffrage, equal access to education and freedom of speech. Eventually this movement had major successes in transforming English social institutions.

Utilitarianism fell out of favor in economics for a variety of reasons. Two principal ones had to do with measurement issues and philosophical perspectives about economic science: (1) Economists came to believe that there was no workable way to measure individual utility. (2) In order to achieve the goal of establishing a moral society, utilities needed to be compared for groups of individuals in order to provide an aggregate utility, thus to measure the overall value to society. Ultimately Utilitarians provided no credible way of doing this. Some critics presented arguments that such comparisons were impossible, and that therefore total utility as a measure of overall social value to society was meaningless.

The most important reason for its falling out of favor in Economics is summarized in Robbins' epigraph at the beginning of this section. It was realized that the mathematical development of economic laws and principles derived by the Utilitarians could be developed without assuming psychological underpinnings. From an economic perspective, this produced a cleaner economic theory. Instead of strong forms of measurement by absolute, ratio, or interval scales for individual utility, all that was needed was ordinal comparisons. The rationale for reform that motivated the Philosophical Radicals of the 19th century that maximizing happiness should be the moral standard by which our actions should be measured was largely lost. Of course, as noted by Hicks (1939), utilitarian ideas could be brought back into economics as an add-on:

> . . . this does not mean that if any one has any other ground for supposing that there exists some suitable quantitative measure of utility, or satisfaction, or desiredness, there is anything in the [ordinal-utility] argument to set against it. If one is a utilitarian in philosophy, one has a perfect right to be a utilitarian in one's economics. But if one is not (and few people are utilitarians nowadays), one also has the right to an economics free of utilitarian assumptions. (p. 18).

The theories of measurement available in 19th and early 20th centuries were inadequate for providing a rigorous, scientific foundation for the kinds of utility theory needed in utilitarianism. That is not so now. Rigorous, scientific methods for the measurement of pleasure are possible today due to the development of the new psychophysical methods of measurement discussed in Sections 4 and 7. Section 11 applies them to the measurement of utility. The comparison of utilities across people was a serious problem for Utilitarianism. Section 11 shows that under some social-economic conditions, a form of interpersonal comparison is possible that leads to a utilitarian group utility. Other results show that group utility comparisons and maximization is achievable without strong interpersonal comparisons.

This article is divided into two parts. Part 1 presents the psychology and measurement theory used by psychologists in modern psychophysics. Part of this psychology in a somewhat confused state was already known to 19th century Utilitarians. This is Fechner's theory of sensory discrimination. The confusion was largely due to Fechner's exposition, which often did not adequately capture his mathematical methods. The psychophysicist Stevens in 1950s and 60s provided alternative methods for intensity measurement that greatly impacted psychology and the social sciences, but had serious flaws as a scientific theory of measurement. It was not adopted by economists.

Part 1 discusses these and more recent forms of psychophysical measurement. Part 2 uses them to develop a scientifically based utility theory consistent with many of the objectives of the Utilitarians Bentham and Edgeworth. We use it to compare utility across individuals.

1.1 Measurement scales and meaningfulness

> *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 raised fundamental questions about measurement. But a rigorous and nuanced theory of meaningful measurement was not developed until the 20th century, starting with Scott & Suppes (1958) article, "Foundational aspects of theories of measurement", and later elaborated by the monographs of Krantz, Luce, Suppes, & Tversky (1971, 1990) *Foundations of Measurement, Vol. I*, Suppes, Krantz, Luce, & Tversky (1990), *Vol. II,* Luce, Suppes, Krantz, Tversky (1990), *Vol. III*, and by 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 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. One 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 $x \to rx + s$, with r positive and 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 any one of its representations, for example, the Kelvin representation.) More empirical structure corresponds to smaller classes of transformations preserving it. The smallest class consists of the identity transformation, and its scale type is said to be *absolute. 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. Under this definition of "meaningfulness", all numerical relationships are meaningful for an absolute scale.

There are scales of measurement intermediate in strength between those envisioned by Anonymous 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.[1] 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".[2] 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.

Part 1 Psychophysics

2 Fechnerian Measurement

The Tower of Babel was never finished because the workers could not reach an understanding on how they should build it; my psychophysical edifice will stand

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

[2]Nozick (1974).

because the workers will never agree on how to tear it down.

Fechner. Über die psychischen Massprincipien und das Weber'sche Gesetz, *Philosophische Studien, 1887, 4,* p. 215.

Bentham had two approaches to the measurement of utility: Measuring the amount of pleasure-pain in an hedonic episode as an integral of hedonic intensities spread across the episode; and measuring it in terms of the maximum number of just noticeable differences experienced within the episode. Kahneman, Walker, & Sarin (1997) and Sarin & Walker (1997) provided Bentham's integral approach with a modern mathematical foundation. Section 9 uses a different but related approach from the theory of measurement to provide a version of Bentham's approach with a foundation similar to the those used for the measurement of basic physical quantities like length, mass, etc. Sections 3 and 4 provide modern treatments of the just-noticeable difference approach. 20th century economic positivists and others took both the integral and just-noticeable difference approaches to be impractical for applied economic application. Also, at the time, neither had the scientific rigor demanded by the positivists for the establishment of an economic science.

Psychophysics distinguishes two general types of measurement methods. *Direct measurement* that measures by subjective assignments of numbers or numerical relations to physical stimuli, and *indirect measurement* that measures through objective observations of behavior.

Consider the case where x is a stimulus from a physical dimension, for example, a light of a particular wavelength and physical intensity. Then a method of direct method championed by Stevens (1946, 1950) would ask the subject to produce a light y that 2 times as bright as x, and with other similar numerical judgments—e.g., produce a light that is 3 times as bright as x, 4 times as bright as x, etc. From judgments like these, Stevens constructed a psychological representation for brightness ψ that measured the subject's subjective intensity of brightness ψ through equations like $\psi(y) = 2\psi(x)$, etc. Fechner (1860) championed a different and indirect approach to brightness. He would provide the subject with two physical light stimuli x and z and force him to make a choice of which was brighter. If x were chosen 75% of the time as being brighter than z, then x is said to be 1 "just noticeable difference" (jnd) brighter than z. For any two stimuli, v and w, with w less physically intense than v, the subjective difference in brightness between v and w is defined to be the largest number of jnds between v and w.

Fechner believed that direct measurement was unscientific. Stevens thought that indirect measurement produced poorer measurements of subjective intensity than his direct measurement methods. The controversy continues today about the merits of the two approaches.

As is implicit in Fechner's epigraph at the beginning of this section, Fechner's psychophysical methodology brought about enormous controversy. This article skips a discussion of the controversies and instead focuses the discussion on his famous law. Fechner provided different arguments for this law. We use the one that fits best in with modern psychophysics and this article's approach to utilitarian utility theory. Dzhafarov & Colonius (2011) write,

> Fechner's writing is often less than clear and open to conflicting interpretations. This article is more of a reconstruction than a review or historical analysis: we try to reconstruct the logic of the Fechnerian approach, and we do this using the language acceptable in modern psychophysics rather than Fechner's own words. Our reconstruction, however, is not an alternative reality, a substitution of what ought to have been said for what has been said. We ascribe to Fechner's theory only the positions that are unequivocally contained in Fechner's texts or can be plausibly inferred from them. Thus, it is a fact that the *Elemente* contains two derivations of Fechner's law, one of which is based on presenting a certain principle (which we call the "W-principle") as a Cauchy-type functional equation. It is a fact that neither derivation makes use of JNDs; therefore, neither derivation is based on Weber's law or the postulated subjective equality of JNDs (known today as "Fechner's postulate"). It is a fact that the counting of just-noticeable increments leading from one stimulus to another as a procedure for measuring subjective difference between them is understood by Fechner as an approximation only, justified if Weber's fraction is sufficiently small. *(p. 2)*

This article follows Dzhafarov's and Colonius's reconstruction. We use a formulation of what is known as "Fechner's Law" that Dzhafarov and Colonius calls "Fechner's forgotten derivation", which appears in Chapter 17 of Fechner (1860).

This reconstruction starts with defining for a physical dimension the *absolute threshold* as the lowest level of the physical stimulus that is detectable by the subject. The absolute threshold is denoted by the symbol o. In Volume 2 of *Elemente*, which has not been translated into English, Fechner formulates a concept of *difference sensation* as follows:

> The sensation of difference between two stimuli is the same as the increment in sensation magnitude from the lesser to the greater of the two stimuli (cf. Elemente, p. 85 of vol. 2). *(Dzhafarov & Colonius, 2011, p. 3)*

This difference, D, is assumed by Fechner to have the following three properties for the physical measurements $\varphi(a)$, $\varphi(b)$, and $\varphi(c)$ of the physical stimuli a, b, c such that $\varphi(a) \leq \varphi(b) \leq \varphi(c)$:

(1) $D(\varphi(a), \varphi(b)) = 0$ iff $\varphi(a) = \varphi(b)$.

(2) $D(\varphi(a), \varphi(b)) = D(\varphi(b), \varphi(a))$.

(3) $D(\varphi(a), \varphi(c)) = D(\varphi(a), \varphi(b)) + D(\varphi(b), \varphi(c))$.

Dzhafarov & Colonius, 2011, p. 4, says the following about (3): "This additivity property is central for Fechner's theory, as he repeatedly states when discussing the notion of measurement (e.g., Elemente, pp. 56, 60 of vol. 1, and Chapter 20 in vol. 2)."

W-Principle: There exists a nonnegative function F such that for all physical measurements $\varphi(o) \leq \varphi(a) \leq \varphi(b)$ from a dimension,

$$D(\varphi(a), \varphi(b)) = F\left(\frac{\varphi(a)}{\varphi(b)}\right) . \tag{1}$$

Equation 1 together with $D(\varphi(a), \varphi(c)) = D(\varphi(a), \varphi(b)) + D(\varphi(b), \varphi(c))$ gives rise to the functional equation,

$$F\left(\frac{\varphi(b)}{\varphi(a)}\right) + F\left(\frac{\varphi(c)}{\varphi(b)}\right) = F\left(\frac{\varphi(c)}{\varphi(a)}\right) , \tag{2}$$

which by well-known techniques functional equations, which were available to Fechner, the solutions to Equation 2 are nonnegative F and positive k such that

$$F(x) = k \log(x), \text{ where } l \leq x \,,$$

which by Equations 1 and 2 yields,

$$\textit{Fechner's difference formula:}\ D(\varphi(a), \varphi(b)) = k \log\left(\frac{\varphi(b)}{\varphi(a)}\right) . \tag{3}$$

3 JND Measurement

The above derivation of Fechner's difference formula used neither the concept of just-noticeable difference nor what is called "Weber's Law". The following is a formulation of Weber's Law:

There exists a nonnegative function F and a positive constant c such that for all physical measurements φ on a dimension, if $\varphi(o) \leq \varphi(a) < \varphi(b)$, where o is absolute threshold and b is just-noticeably different from a, that is, b 1 jnd from a, then

$$\textit{Weber's Law:}\ \frac{\varphi(b)}{\varphi(a)} = 1 + c \,, \tag{4}$$

where c is *Weber's constant* given by the formula,

$$\frac{\varphi(b) - \varphi(a)}{\varphi(b)} = c \,.$$

We say that a function θ on the reals is a *jnd representation of size* d if and only if there exists a positive real number d such that for all a on the dimension of stimuli,

$$\theta[\varphi(a)] - \varphi(a) = d \,.$$

In this case, d is said to be "1 jnd".

Assume Fechner's difference formula and Weber's Law. Then, on the sensation difference measurement representation D, it follows from Equation 3 that the difference of 1 jnd has constant value $k \log(1+c)$ *everywhere on the* D *representation.* However, a jnd representation for a situation satisfying Weber's Law is not sufficient for having the Fechner difference formula hold.[3]

4 Semiorders and Theshold Structures

...when it is explicitly brought to my attention that I have shown a preference for f *as compared with* g*, for* g *as compared with* h*, and for* h *as compared with* f*, I feel uncomfortable in much the same way that I do when it is brought to my attention that some of my beliefs are logically contradictory.*

Leonard Savage, *The Foundations of Statistics,* 1954, p. 211.

A modern, axiomatic approach to jnds was initiated by Luce (1954). He defined an ordering $x \prec y$ on stimuli that was meant to capture the idea that y was at least 1 jnd more than x in subjective intensity. He presented axioms about $\prec$ where each corresponded to a simple experiment, and showed that if his axioms held then there was a measurement function Ψ on the stimuli such that

$$x \prec y \text{ iff } \Psi(x) + 1 \leq \Psi(y)\,,$$

that is, the psychological intensity $\Psi(y)$ of y is at least 1 jnd (+1) more intense than the psychological intensity $\Psi(x)$ of x.

In psychophysics, the set of stimuli is usually a physical continuum measured physically by a function φ onto the ordered set of positive reals. Luce's axiomatization was for a finite set of stimuli. His method of proof did not extend to continua. Narens (1994) extended it so that it applied to continua. The following is equivalent to Luce's axiomatization.

$\prec$ is said to be a *semiorder* if and only if X is a nonempty set, $\prec$ is a binary relation on X, and the following three axioms are true for all w, x, y, and z in X.

1. Not $x \prec x$.
2. If $w \prec x$ and $y \prec z$, then $w \prec z$ or $y \prec x$.
3. If $w \prec x$ and $x \prec y$, then $w \prec z$ or $z \prec y$.

[3]Representations other than θ that have exactly the same jnd structure as θ can be constructed from θ as follows: Let g be any strictly monotonic function from absolute threshold $\varphi(o)$ to $\theta[\varphi(o)]$ such that $g[\varphi(o)] = \varphi(o)$, $g(\theta[\varphi(o)]) = \theta[\varphi(o)]$, and for x in the interval $(\varphi[o], \theta[\varphi(o)])$, $g(x) \neq x$. For each integer n, if y is in the interval $[n \cdot \varphi(o),\, n \cdot \theta(\varphi(o)]$, let $\sigma(y) = ng(x)$, where $x - n$ is in the interval $[\varphi(o),\, \theta[\varphi(o)]]$. Then σ is another jnd representation that is different from θ but agrees everywhere with θ on jnd differences.

These axioms, which are observable through subjects' behavior, are simple and testable. Parts of axioms 2 and 3 (e.g., "$w \prec x$") are about reports of the subject's subjective experience. However, the entire set of axioms combine these reports in a manner that is not directly experienced by the subject. That is, these axioms, which are observable to the experimenter, are not directly experienceable by the subject, unless they are presented in their entirety to a subject as a four stimuli trial, for example: Present the previous results $w \prec x$ followed by $y \prec z$ to the subject, and ask, "Which of the following is true: $w \prec z$ or $y \prec x$?" By looking at behavior instead of subjective experience, stronger discrimination can be found. In particular, the semiorder $\prec$ can be used to define a weak ordering $\precsim_*$ for comparing subjective intensities of stimuli in X.

Define $\sim$ on X by: For all x and y in X,

$$x \sim y \text{ iff } x \prec y \text{ and } y \prec x\,.$$

Then, except for degenerate cases, $\sim$ is *not* an equivalence relation, that is, generally there are x, y, and z in X such that $x \sim y$, $y \sim z$, but not $z \sim x$—illustrating Savage's lament in the epigraph at this section's beginning.

Let $\prec$ be a semiorder on X. Define $\precsim_*$ as follows: For all x and y in X,

$$x \precsim_* y \text{ iff for all } z[(\text{if } y \prec z \text{ then } x \prec z) \text{ and } (\text{if } z \prec x \text{ then } z \prec y]\,.$$

$\precsim_*$ is called the *weak order induced by* $\prec$. It not difficult to show that $\precsim_*$ is a weak order, that is, is a transitive and connected relation (Luce, 1954).

Luce (1954) showed the following theorem.

Theorem 1 *(Luce's Representation Theorem)* Suppose X is a finite set of stimuli and $\prec$ is a semiorder on X. Then there exists a function Ψ from X into the real numbers such that for all x and y in X,

$$x \prec y \text{ iff } \Psi(x) + 1 < \Psi(y)\,.$$

If we say for real numbers r and s, $r < s$, that s is 1 jnd more than r if and only if $r + 1 = s$, then $x \prec y$ iff the measurement $\Psi(y)$ of y is at least 1 jnd more than the measurement of x, $\Psi(x)$.

While experiments are about finite sets of stimuli X, most of economic and psychophysical theory are about infinite X on a weakly ordered continuum. Unfortunately the method of proof of the Luce Representation Theorem does not extend to continua of stimuli. Narens (1994) came up with an approach to handle this.

$\langle X, \precsim' \rangle$ is said to be a *weakly ordered continuum* if and only if $\precsim'$ is a weak ordering on X and there is a strictly order preserving function φ from X onto the reals, $\mathbb{R}^+$, that is, for all x and y if X,

$$x \precsim y \text{ iff } \varphi(x) \le \varphi(y)\,.$$

$\langle X, \precsim', T \rangle$ is said to be a *continuous threshold structure* if and only if $\langle X, \precsim' \rangle$ is a weakly ordered continuum and T is a function from X onto X such that for each x in X, $x \prec' T(x)$. T is called the *threshold function* for $\langle X, \precsim', T \rangle$.

The following two theorems follow from the definitions of "semiorder", "weak order induced by $\prec$", and "continuous threshold structure".

Theorem 2 Suppose $\langle X, \precsim', T\rangle$ is a continuous threshold structure. Define $\prec$ on X as follows: for all x and y in X,

$$x \prec y \text{ iff } T(x) \prec' y\,.$$

Then $\langle X, \prec\rangle$ is a semiorder and $\precsim'$ is the weak order induced by $\prec$.

Suppose $\langle X, \prec\rangle$ is a semiorder. Then $\langle X, \prec\rangle$ is said to be *continuous* if and only if $\langle X, \precsim_\star\rangle$ is a continuum.

Theorem 3 *Suppose $\langle X, \prec\rangle$ is a continuous semiorder and $\precsim_\star$ is the weak order induced by $\prec$. Define the function T on X as follows. For all x and y in X, $T(x) = y$ if and only if for all z in X,*

*(*i*) if $y \prec_\star z$ then $x \prec z$, and*

*(*ii*) if $z \prec_\star y$ then $x \not\prec z$.*

Then $\langle X, \precsim_\star, T\rangle$ is a continuous threshold structure.

Narens (1994) showed the following representation theorem.

Theorem 4 *(Threshold Representation Theorem)* Suppose $\langle X, \preceq', T\rangle$ is a continuous threshold structure. Then there exists a function Ψ from $\langle X, \preceq'\rangle$ onto $\langle \mathbb{R}, \leq\rangle$ such that for all x and y in X,

$$T(x) \prec' y \text{ iff } \Psi(x) + 1 < \Psi(y)\,.$$

5 Plateau

Fechner's formula leads to this consequence that, when the overall illumination increases, the differences in sensation remain constant; it seemed to me more rational, in order to explain the invariance of the general effect of the picture, to postulate a priori *the constancy of the ratios and not the differences of the sensations.*

Plateau (1872), pp. 382–383. Translated in Falmagne (1985), pg. 318, with emphasis on "a priori" added.

Contrary to Fechner, Plateau (1872) concluded that the psychophysical function F obeyed the power law $f(x) = ax^b$, where a, b, and x are positive, a and b are constants and x is a variable. He based his argument on invariance considerations involving a subjective judgment of bisecting two visually presented stimuli. This is direct measurement, because Plateau assumed that subjects produced a stimulus that was subjectively "midway the other two in terms of grayness".

His experiment consisted of providing eight artists with two disks—one painted black and the other white—with the instruction to paint a gray disk midway between them. The eight midway disks returned to Plateau were almost identical. This, despite presumed illumination differences under which they were painted. He made the following assumption: Each artist mixed his paint to obtain a gray such that the ratio of the subjective intensity of white to gray equaled the subjective intensity of gray to black. Considering the black and white disks as extreme examples of gray ones and assuming a similar result for any pair of gray disks then yields the following functional equation,

$$\frac{\psi(d)}{\psi(m)} = \frac{\psi(m)}{\psi(e)} \, ,$$

where d and e are the disks provided for midway judgment, m is the midway disk painted by the artist, and ψ is a function that measures the subjective intensity of grayness. It is a known physical fact that the *ratios* of the usual physical measurements of the light for such disks do not vary with illumination. Because of this, Plateau reasoned that such ratios would be the same in each artist's studio. Combing this with his experimental results, Plateau concluded that the following law held that we call the *Preserved Midway Ratio Law:* For all gray disks d and e,

$$\frac{\varphi(d)}{\varphi(m)} = \frac{\varphi(m)}{\varphi(e)} \text{ iff } \frac{\psi(d)}{\psi(m)} = \frac{\psi(m)}{\psi(e)} \, , \tag{5}$$

where m is the gray disk produced midway between d and e, φ measures the physical intensity of grays, and ψ measures subjective grayness. The functional equation solution to Equation 5 is the *psychophysical power law,*

$$\psi = r\varphi^s$$

for some positive r and s.

6 Stevens

For seven years a committee of the British Association for the Advancement of Science debated the problem of measurement. Appointed in 1932 to represent Section A (Mathematical and Physical Sciences) and Section J (Psychology), the committee was instructed to consider and report upon the possibility of "quantitative estimates of sensory events"—meaning simply: Is it possible to measure human sensation? Deliberation led only to disagreement, mainly about what is meant about the term measurement. An interim report in 1938 found one member complaining that his colleagues "came out by the same door as they went in," and in order to have another try at agreement, the committee begged to be continued for another year.

For its final report (1940) the committee chose a common bone for its contentions, directing its arguments at a concrete example of a sensory scale. This *was the Sone scale of loudness (S. S. Stevens and H. Davis.* Hearing. *New York: Wiley, 1938), which purports to measure the subjective magnitude of an auditory sensation against a scale having the formal properties of basic scales, such as those used to measure length and weight. Again 19 members of the committee came out by the routes they entered, and their view ranged widely between two extremes. One member submitted "that any law purporting to express a quantitative relation between sensation intensity and stimulus intensity is not merely false but is in fact meaningless unless and until a meaning can be given to the concept of addition as applied to sensation" (Final Report, p. 245).*

Stanley Stevens, On the theory of scales of measurement, *Science, 103,* 1946, pp. 677–680.

Steven's response to the Final Report was to develop a new system of measurement based on rules for assigning numbers to objects instead of the standard one at the time that depended on properties of a qualitative operation corresponding to adding a pair of assigned numbers. For Stevens, the rule could be any well-defined systematic method of assignment. What was important is specifying all the assignments that could be made by specifying the rule in a particular situation. These form a scale for measuring the situation. Stevens added a condition that statistical conclusions drawn about the situation had to be invariant under the assignments from the scale. He called this condition *meaningfulness.* He also provided an algebraic characterization of the kinds of scales that he found in use in science at the time. A main criticism of his system at the time was that his concept of "rule" was too general or vague. He provided the following controversial examples of magnitude estimation and production as uses of rules for psychophysical measurement. They, and related methods became widely used—and continue to be widely used—in the behavioral and social sciences.

Magnitude estimation has the subject provide a numerical estimate of the ratios of two stimuli from a dimension, for example, the ratio of the brightnesses of two lights. In *magnitude production,* the experimenter presents a stimulus x, called the *modulus,* from a dimension and a number word n and asks him to find the y from the dimension that is "n times x." In both paradigms, the physical measures of the stimuli are recorded as well as the number the number word names.[4] Stevens viewed his form of measurement operationally as a relation between stimulus and response. Thus, in his view, it was not a form of direct measurement. But Fechner would certainly had considered it as such. And we will do so as well. The slight of hand—or if one prefers to call it "a change of philosophy of science"—of the experimenter recording the response of a number word as the number denoted by word, does not change the fundamental problem of interest to psychology: the relationship of the physical

[4]Stevens either confused numerals (number words) with numbers or incorrectly thought that science could work with numerals replacing numbers without providing an empirically based rationale concerning the operations such as addition, multiplication, etc., required in mathematical-scientific applications. Also he provided no test of the consistency of such assignments with how arithmetic is used in science and statistics—see Section 7 below.

world to mental representations of it and how these are mediated through an intervening physiology.

Stevens, colleagues, and many other researchers conducted a multitude of magnitude estimation and production experiments that empirically demonstrated the validity of the power law. Many outside of the behavioral and social science communities, as well as some of the more mathematically and philosophically sophisticated researchers within those communities, were suspicious of the measurement methods rooted in the magnitude experiments. They, however, respected the wide body of consistently related and replicated empirical results that were generated its methods in psychophysical studies. The 1960s and 70s developed new mathematical techniques of measurement. The empirical results generated by Stevens' methods, although believed to be acquired through non-rigorous methods, continued to be respected by those psychological measurement theorists.

Contemporaneous with Stevens was an approach much like Plateau's that was based on a functional equation argument. It assumed a physical measuring function φ on a physical dimension of stimuli and a psychological measuring function ψ on the stimuli such that for each quadruple of stimuli x, y, w, z of stimuli,

$$\textit{Equal Ratios Law:} \text{ If } \frac{\psi(x)}{\psi(y)} = \frac{\psi(w)}{\psi(z)} \text{ then } \frac{\varphi(x)}{\varphi(y)} = \frac{\varphi(w)}{\varphi(z)} \,. \tag{6}$$

The well-known functional equation solution to Equation 6 is the psychophysical power law,

$$\psi = r\varphi^s \,, \tag{7}$$

for some positive r and s. Note that the Equal Ratios Law does not specify how ψ is to be determined empirically.

Stevens generally used empirical methods for supporting the power law, i.e., by interpreting veridically the number words used or produced in the experiment, and appropriately plotting how the physical measures of the stimuli were related to the interpretations to obtain an approximation of the power law. The use of the Equal Ratio Law gets the power law very differently. Consider the case of magnitude production for brightness. Give the subject lights x, y, and w and have her adjust lights until she finds a light z satisfying,

$$\frac{\psi(x)}{\psi(y)} = \frac{\psi(w)}{\psi(z)} \,. \tag{8}$$

The experimenter can easily fine $\varphi(x)$, $\varphi(y)$, and $\varphi(w)$ because they are physical measurements. Equation 8 then provides an empirical test for the Equal Ratios Law, because it specifies the physically measured value $\varphi(z)$ must have for Equation 6 to hold.

7 Making Direct Measurement Indirect

A primary aim of measurement is to provide a means of convenient computation. Practical control or prediction of empirical phenomena requires that unified, widely applicable methods of analyzing the important relationships between the phenomena be developed. Imbedding the discovered relations in various numerical relational systems is the most important such unifying method that has yet been found.

Dana Scott and Patrick Suppes, Foundational aspects of theories of measurement, 1958. *Journal of Symbolic Logic, 23,* pp. 113–128.

Plateau had his subjects paint a gray midway M between between two given grays. He assumed direct measurement—that is, that the subject painted M according to some scientific understanding of "midway". But, What is this understanding? Can it be observed from the subject's behavior? If it could, then the "direct" in "direct measurement" could be replaced by "indirect". The following algebraic characterization and theorem shows how to accomplish this.

$\langle X, \preceq, \otimes\rangle$ is said to be a *continuous bisection structure with bisection operation* $\otimes$ if and only if $\otimes$ is a binary operation on X and the following six conditions hold for all w, x, y, and z in X:

1. *Continuum:* $\langle X, \preceq\rangle$ is a continuum, that is, there is an isomorphism of $\langle X, \preceq\rangle$ onto $\langle \mathbb{R}, \leq\rangle$ or $\langle \mathbb{R}^+, \leq\rangle$.
2. *Idempotence:* $w \otimes w = w$.
3. *Commutativity:* $x \otimes y = y \otimes x$.
4. *Solvability:* There exists v in X such that $u \otimes x = v \otimes y$.
5. *Monotonicity:* $x \preceq y$ if and only if $x \otimes z \preceq y \otimes z$.
6. *Bisymmetry:* $(w \otimes x) \otimes (y \otimes z) = (w \otimes y) \otimes (x \otimes z)$.

Theorem 5 *Suppose* $\mathfrak{X} = \langle X, \preceq, \otimes\rangle$ *is a continuous bisection structure. Then* $\mathfrak{X}$ *and* $\mathfrak{N} = \langle \mathbb{R}^+, \leq, \otimes^\star\rangle$ *are isomorphic, where* $\otimes^\star$ *is the binary operation on* $\mathbb{R}$ *such that for all* r *and* s *in* $\mathbb{R}^+$*,*

$$r \otimes^\star s = (r \cdot s)^{\frac{1}{2}} .$$

Proof. Theorem 10 of Section 6.9 of Krantz, Luce, Suppes, and Tversky (1971) and Theorem 2.1 of Luce and Narens (1985).[5]

[5] The first four axioms of Theorem 10 of Section 6.9 of Krantz et al. (1971) easily follow from the Conditions 1 to 6. The remaining axiom (the Archimedean axiom) follows from Theorem 2.1 of Luce and Narens (1985).

In Plateau's demonstration, let $u \preceq x$ stand for "u is less than or the same amount as gray than x" and $u \otimes x$ stand for "the gray midway between u and x". Then if the existence of arbitrary large stimuli is assumed (a common mathematical idealization), then all the conditions for a continuous bisection structure hold, except possibly for bisymmetry,

$$(u \otimes x) \otimes (y \otimes z) = (u \otimes y) \otimes (x \otimes z)\,. \tag{9}$$

Equation 9 corresponds to the experiment: the midway gray disk produced by the left side of Equation 9 is identical in the appearance of grayness to the one produced by the right side of the equation. Thus bisymmetry is the key experimental condition to test in order to see if the subject's "midway" gray has the correct mathematical property of being equal ratios from the grays it is between. If so, isomorphism of "midway" has been reach with the geometric mean, and the bisection operation $\otimes$ is indirect.

Producing an indirect version of Stevens magnitude estimation and production is more complicated and subtile. Stevens' rules for assigning numbers to magnitude estimates and productions are clear, but it is not clear what else is needed if a clear algebraic formulation is to be given. Narens (1996) produces a theory that does this. It will be stated for magnitude production; an analogous version can be given for magnitude estimation.

Narens (1996) isolates two principles that he believes are inherent in Stevens' direct methods of measurement. Although these principles are not stated by Stevens, they, according to Narens they are implied by,

(i) his data collection methods,

(ii) his methods of statistical analysis, and

(iii) the conclusions he drew from his experimental research.

Narens (1996) reformulates Stevens' theory into the following two principles:

> Principle 1: $\mathcal{S}$ is a ratio scale family that adequately measures the observer's subjective intensity of stimuli in X.
>
> Principal 2: For each t in X, there is a function ψ_t in $\mathcal{S}$ such that (i) $\psi_t(t) = 1$, and (ii) for each x in X, if the subject's behavior is that she produces x as the response to "p-times t", then $\psi_t(x) = p$.

Then Theorem 6 below shows that Principles 1 and 2 imply the "multiplicative property" defined below.

Throughout the rest of this section, the following notation is used for all t and x in X and positive integers p: $f_p(t) = x$ if and only if x is the subject's response to "Find the stimulus x that is p-times in intensity to stimulus t". f_p is called a *magnitude production function.* Note that in terms of ψ_t in Principles 1 and 2,

$$f_p(t) = x \text{ iff } \psi_t(x) = p\,. \tag{10}$$

For some purposes the notation "f_p" is more revealing, and for others the notation "ψ_t" is more revealing.

Definition 1 *Multiplicative Property:* For all for t in X and all positive integers p and q,

$$f_q[f_p(t)] = f_{q \cdot p}(t)\,.$$

Theorem 6 *Suppose $\mathcal{S}$ is a scale of representations from X onto $\mathcal{R}^+$ satisfying Principles 1 and 2. Then the Multiplicative Property holds.*

Proof. Let $\psi_t(x) = p$, $\psi_x(y) = q$, and $\psi_t(y) = r$. By Principles 1 and 2, let s be a positive real such that

$$\psi_x = s \cdot \psi_t\,.$$

Then

$$1 = \psi_x(x) = s \cdot \psi_t = s \cdot p$$

that is,

$$s = \frac{1}{p}\,.$$

Thus,

$$q = \psi_x(y) = s \cdot \psi_t(y) = \frac{1}{p} \cdot \psi_t(y) = \frac{1}{p} \cdot r\,,$$

in other words, $r = p \cdot q$. In terms of f_p and f_q this yields,

$$f_q[f_p(t)] = f_{q \cdot p}(t)\,.$$

The multiplicative property corresponds to a simple experiment. It has been tested many times in several sensory domains by several laboratories and has been found to fail for loudness and brightness. (See Luce, Steingrimsson, Narens, 2010, for references.) For judgments of distance ratios, it holds in about half of the subjects (Bernasconia, Choiratb, & Seria, 2008). We interpret these results to mean that Stevens' empirical data was consistent with his method of measurement and the assumptions underlying his theory only held because he did not collect relevant data that would lead to their refutation. For example, the Multiplicative Property implies producing a sound x that is twice as loud as t, and the sound y that three times as loud x, should yield a sound z that is six times as loud as t. Empirical results of Ellermeier & Faulhammer (2000) and others show that such z are produced more like twelve times as loud as opposed to six—a very large discrepancy. (Zimmer, 2005, found similar results using $\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{6}$, and $\frac{1}{12}$.) Stevens did not did not conduct simple tests like this.

Narens (1996) predicted that the Multiplicative Property would likely fail and suggested using a weaker principle called the "Commutativity Property", which is implied by the Multiplicative Property, might be more appropriate for evaluating magnitude production judgments.

Commutative Property: For all for t in X, and all positive integers p, q,

$$f_q[f_p(t)] = f_p[f_q(t)]\,.$$

Narens (1996) showed the following theorem.

Theorem 7 *Assume that commutativity property. Then there exists a ratio scale $\mathcal{S}$ of representations on X and a function W from $\mathbb{R}^+$ onto $\mathbb{R}^+$ such that for all x in X, positive integers p, and φ in $\mathcal{S}$,*

$$\varphi[f_p(x)] = W(p) \cdot \varphi(x)\,. \tag{11}$$

Of course, when $W(p) = p$, Equation 11 reduces to Stevens method magnitude production.

Equation 11 illustrates the kind of structure preserving mapping suggested in the Scott-Suppes epigraph. More formally, call the structure $\langle X, \preceq, f_p \rangle_{p \in I^+}$ a *magnitude production structure,* where $\langle X, \preceq \rangle$ is a continuum and the f_p are magnitude production functions. Narens (1996) gave an indirect measurement account of magnitude production structures that resulted in Equation 11 with $W(p) = p$, that is, provided an indirect account of the results that would be obtained by Stevens method of magnitude production applied in an idealized setting. Most of the assumptions made were simple, obvious ordinal ones, e.g., $t \preceq f_p(t)$, or necessary for mapping onto the positive real numbers. The main questionable assumptions needed empirical justification were the following two:

(1) For all t in X, $f_1(t) = t$.

(2) The Multiplicative Property.

The resulting theorem is that *there exists a ratio scale $\mathcal{S}$ such that Equation 11 holds with $W(p) = p$.* Replacing the Multiplicative Property with the Commutative Property yields the more general theorem, *there exists a ratio scale $\mathcal{S}$ such that Equation 11 holds.*

Surprisingly, (1) for all t in X, $f_1(t) = t$, fails in loudness experiments when explicitly tested for (Steingrimsson & Luce, 2012). However, it holds in brightness experiments. The apparent reason is that in loudness the two stimuli must be given successively for evaluation, whereas in brightness they can be given simultaneously. Paradigms using commutativity can be altered to take into account the failure of (1) (e.g., Luce, et al., 2010) at the cost of having a more complicated formula than Equation 11 that has an additional constant to be estimated from data. A model comparison test involving distance and area judgments was conducted (Bernasconia, Choiratb, & Seria, 2008), involving Equation 11, the more complicated Commutative Property model, and Stevens' method. All three statistically fit the data. Equation 11 and the more complicated Commutative Property Model were better fits than Stevens' method.

In summary, the two main direct methods used in psychophysics—bisection and magnitude estimation and production—can be given indirect formulations on a par with the measurement models used in physics. The data collected in Stevens' magnitude paradigms fit theories based on the Commutativity Property better than Stevens' theory.

8 Torgerson's Cojecture

Several years ago, Garner (1954) tried to get subjects first to set a variable stimulus between two standard stimuli so that the successive differences were equal, and second, to set the variable so that the successive ratios were equal. That is, first, so that $V - S_1 = S_2 - V$, and second, so that $V/S_1 = S_2/V$. For most of his subjects—thirteen out of eighteen—the value set for the variable was the same in the two conditions: Equal subjective intervals were also equal subjective ratios.

Warren Torgerson, 1961, p. 204.

Torgerson (1961) discussed the above experiment of Garner and several others involving direct judgments of subjective intensity and concluded,

> These results are all consistent with the notion that the subject perceives only a single quantitative relation between stimuli. When this relation is interpreted as either a psychological distance or a psychological ratio, it can be shown that the subjective magnitudes obey the properties of the corresponding commutative group—the addition group for the distance interpretation and the multiplication group the ratio interpretation. *(p. 205)*

This conclusion is often called *Torgerson's Conjecture.* It is controversial. There are many experiments supporting it and several contradicting it. Also, many of the studies in support are based on direct measurement techniques, making the results suspicious to many researchers. However, later studies eliminated this concern, (eg., Birnbaum & Mellers, 1978). Despite the controversy, it is striking that there are many carefully performed studies—several with seasoned, sophisticated, psychophysical subjects—that demonstrate the Conjecture. This calls into question the conclusions drawn from many commonly used psychological, social, and behavioral techniques of measurement. This extends as well to the economic concept of "utility" or "value". A somewhat similar observation involving the necessity of comparing pairs as Torgerson's, but based on the exchangeability of economic commodities instead of psychology, was put forth by the 19th century Utilitarian philosopher Samuel Baily concerning the measurement of value:

> As we cannot speak of the distance of any object without implying some other object between which and the former this relation exists, so we cannot speak of the value of a commodity, but in reference to another commodity compared with it. A thing cannot be valuable in itself without reference to another thing, any more than a thing can be distant in itself without reference to another thing.

Samuel Baily, *A Critical Dissertation on Value,* 1825, p. 5.

Part 2 Individual and Group Utilities

9 Bentham

Hence to know what men will do, to tell what they should do, or to value what they have done, one must be able to measure varying "lots" of pleasure or pain. How are such measurements to be made?

Bentham, *Introduction to the Principles of Morals and Legislation, Works, vol. i,* pp. I

Bentham wanted to put the social and behavioral sciences and law on a scientific footing similar to the physics of his day. In order to accomplish this, he realized that this has to be done in such a way that the basic concepts had to have mathematical representations so that underlying mathematical laws could be formulated. He used utility as his basic concept, which he interpreted as the measurement of the hedonic dimension of "pleasure and pain". He sought various means to quantify pleasure-pain. But, unlike many subsequent Utilitarians, he realized that there were grave difficulties in carrying this out. In his scattered writings, he wrote about them. A major part of the problem for him was the lack at the time of adequate, abstract theories of measurement to draw upon and the lack of relevant psychological and economic concepts. He had to create his own concepts. He tried, and from the point of view of this article, accomplished a lot: As results presented here will show, many of his ideas will generate, when using modern ideas from modern psychology and measurement theory, philosophically rigorous versions of his vision about the quantification pleasure-pain. Of course, some changes in his formulation have to be made. Because of this, we call our reformulations "neo-Benthamite theories" of utility. Two will be given. Both give utilities for hedonic episodes. The first is to treat the utility of an episode as an integral of hedonic intensity over time. The second quantifies the utility of episodes through a famous mathematical result of Hölder (1901) that has been used in physics to measure length, time, mass, and other fundamental physical dimensions.

9.1 Benthamite utility as an Integral

In his early work, Bentham measured pleasure-pain lots in terms of the intensities of sensations and their duration. He realized that there were qualitatively different pleasure-pain sensations. These presented measurement difficulties when considered in combination, and similarly for measuring pleasure-pain for a population of individuals. He originally thought that such difficulties could be largely avoided by summing different pleasure-pain lots, but realized later that this wouldn't work.

His first theory of measuring pleasure-pain was in terms of intensity and duration:

> The unit of intensity is the faintest sensation that can be distinguished to be pleasure or pain; the unit of duration is a moment of

> time. Degrees of intensity and duration are to be counted in whole numbers, as multiples of these units.

The unit of intensity multiplied by the unit of duration produced a unit of pleasure-pain, and thus the multiplication of their degrees produced a measure of pleasure-pain. This idea was more clearly and fully worked by Edgeworth and his concept of a "hedonimeter". 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.

Edgeworth (1887) provided a more complete and mathematical formalization of the hedonimeter. Notice that its "unit of intensity is the faintest sensation" is just what Weber called a "just-noticeable difference".

Sarin & Wakker (1997) shows how Bentham's and Edgeworth's ideas lead to a legitimate measurement system. This system has been used in a new paradigm in psychology called "experienced utility" (e.g., see Kahneman, Wakker, & Sarin, 1997).

The hedonimeter model is just assigning the the utility of an *episode* as the integral $\int_a^b f(t)dt$, where $f(t)$ is the hedonic intensity at time t within the episode and a and b are, respectively, the beginning and end of the episode. Note that time, and therefore the difference in time between a and b, are objective and that hedonic intensity $f(t)$ is subjective. This is similar to the von Neumann-Morgenstern model of expected utility, which is also represented as an integral over prospects that produce outcomes of varying subjective intensities over events with objective probabilities. The important difference is that in the hedonimeter model, the utility of episodes varies in a different manner than the utility of outcomes in the von Neumann-Morgenstern model. For example, hedonimeter models have ratio scale utility for hedonic episodes while the von Neumann-Morgenstern models have interval scale utility for outcomes and prospects. For the hedonimeter, utility 0 is assigned for episodes of 0 duration. It is the limit of episodes having bounded hedonic intensity as duration goes to 0. This value and limit is the same for each measuring function from the ratio scale. For von Neumann-Morgenstern, there is no prospect that is invariant under each measuring function from its interval scale.

9.2 Extensive measurement of hedonic utility

> *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.

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 this subsection ignores probabilistic concerns. 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 subsection looks at 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 employable. 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?".

9.2.1 Hedonic episodes

Throughout this section, $\mathcal{H}$ denotes 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 subsection 9.1 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

$$(F^\frown G)^\frown K \sim F^\frown(G^\frown K)\,. \qquad (12)$$

9.2.2 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 12 that $(J^\frown K)^\frown L \sim J^\frown(K^\frown L)$, and thus that

$$\textit{Associativity of } \oplus\textit{:} \ \ (A \oplus B) \oplus D \ \sim \ A \oplus (B \oplus D)\,. \tag{13}$$

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 13), 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.

Subsection 9.1 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 the Benthamite 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, generally, that

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

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.

10 Money as a Measure of Individual and Group Utility

10.1 Bentham's use of Money as a Proxy for Utility

To avoid the difficulties inherent in summing the measurements non-homogeneous lots of pleasure-pain, the later Bentham decided to use money as a proxy for them:

> If of two pleasures a man, knowing what they are, would as lief enjoy the one as the other, they must be reputed equal. . . . If of two pains a man had as lief escape the one as the other, suc'h two pains must be reputed equal. If of two sensations, a pain and a pleasure, a man had 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,* or, as we may say in this case, *equivalent.*
>
> If then between two pleasures the one produced by the possession of money, the other not, a man had as lief enjoy the one as the other, such pleasures are to be reputed equal. But the pleasure produced by the possession of money, is as the quantity of money that produces it: money is therefore the measure of this pleasure. But the other

> pleasure is equal to this; the other pleasure therefore is as the money that produces this; therefore money is also the measure of that other pleasure. It is the same between pain and pain; as also between pain and pleasure.
>
> . . . If then, speaking of the respective quantities of various pains and pleasures and agreeing in the same propositions concerning them, we would annex the same ideas to those propositions, that is, if we would understand one another, we must make use of some common measure. The only common measure the nature of things affords is money.
>
> I beg a truce here of our man of sentiment and feeling while from necessity, and it is only from necessity, I speak and prompt mankind to speak a mercenary language. . . . Money is the instrument for measuring the quantity of pain or pleasure. Those who are not satisfied with the accuracy of this instrument must find out some other that shall be more accurate, or bid adieu to Politics and Morals. (Quoted from Halévy, vol. I, pp. 410, 412, 414.)

Bentham thus took money as a way of measuring pleasure or happiness. He believed in the diminishing marginal utility of money, that is, an increment of a specific amount of wealth at a higher wealth level will produce a smaller increase in happiness than it would at a lower wealth level:

> [The] quantity of happiness produced by a particle of wealth (each particle being of the same magnitude) will be less and less at every particle; the second will produce less than the first, the third than the second, and so on." ["Pannomial Fragments," *Works, vol. iii,* p. 229] . . . for by high dozes of the exciting matter applied to the organ, its sensibility is in a manner worn out. ["Constitutional Code," *Works, vol. ix,* p. 15.]

Mitchell (1918) provides the following succinct summary of the legacy of Bentham's hedonic calculus.

> The net resultant of all these reflections upon the felicific calculus collected from Bentham's books and papers might be put thus: (1) The intensity of feelings cannot be measured at all; (2) even in the case of a single subject, qualitatively unlike feelings cannot be compared except indirectly through their pecuniary equivalents; (3) the assumption that equal sums of money represent equal sums of pleasure is unsafe except in the case of small quantities; (4) all attempts to compare the feelings of different men involve an assumption contrary to fact. That is a critic's version of admissions wrung from Bentham's text; a disciple's version of his master's triumphs might run: (1) The felicific calculus attains a tolerable degree of precision since all the dimensions of feeling save one [i.e., intensity] can be measured; (2) the calculus can handle the most dissimilar feelings

by expressing them in terms of their monetary equivalents; (3) in the cases which are important by virtue of their frequency, the pleasures produced by two sums of money are as the sums producing them; (4) taken by and large for scientific purposes men are comparable in feeling as in other respects. ... Heat these two versions in the fire of controversy and one has the substantial content of much polemic since Bentham's day. *(p. 172)*

10.2 The use of money for comparing utilities across individuals

The deepest measurement issue for Utilitarianism is the comparability of utilities across individuals. Some Utilitarians recognized this as a serious, foundational problem. For example Jevons (1871) 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*)

Other Utilitarians think differently. For example, Edgeworth (1887) writes,

> Utility, as Professor Jevons says, has two dimensions, intensity and time. The unit in each dimension is the just perceivable [Wundt, *Physiological Psychology*] increment. The implied equation to each other of each *minimum sensible* is a first principle incapable of proof. It resembles the equation to each other of undistinguishable events or cases [Laplace, *Essai–Probabilities, p. 7*] which constitutes the first principle of the mathematical calculus of *belief.* It is doubtless a principle acquired in the course of evolution. The implied equatability of time intensity units, irrespective of distance in time and kind of pleasure, is still imperfectly evolved. Such is the unit of *economical* calculus. *p. 7*

Both Jevons and Edgeworth agree that the intercomparability of utility is incapable of scientific verifiability or falsification, with Edgeworth taking it as a fundamental principle of Utilitarianism. Section 11 will argue that their view on scientific verifiability or falsification is wrong. This will be done by showing if conditions are right, then individuals' utilities can be put on a common scale and compared. This will be done by using measurement techniques of that were developed a century or more after their writings on the subject.

Utilitarians, like Bentham, recognized the difficulties in comparing pleasures within and between persons. Some sought to avoid these difficulties by identifying utility with money. But this approach had it own problems. It gives the same utility of a good for a rich man as to poor man. Bentham realized this and tried to correct for it by taking ratios:

> ... the proportion between pleasure and pleasure is the same as that between sum and sum. So much is strictly true that the ratios between the two pairs of quantities are nearer to that of equality than to any other ratio that can be assigned. Men will therefore stand a better chance of being right by supposing them equal than by supposing them to be any otherwise than equal. ...
>
> Speaking then in general, we may therefore truly say, that in small quantities the pleasures produced by two sums are as the sums producing them. *Hallévy, vol. I, pp. 408, 410.*

There are subtleties in using money as a proxy for utility. A few were noted by Marshall and his protege Pigou. For the material in this section, the most important one involves a distinction between satisfaction and desire.

> [the] money which a person is prepared to offer for a thing measures directly, not the satisfaction he will get from the thing, but the intensity of his desire for it ... The substantial point is that we are entitled to use the comparative amounts of money which a person is prepared to offer for two different things as a test of the comparative satisfactions which these things will yield to him, only on condition that the ratio between the intensities of desire that he feels for the two is equal to the ratio between the amounts of satisfaction which their possession will yield to him. *Pigou (1920), p. 23*

Like Bentham, Pigou takes ratios of monies of goods as a proxy for their ratio of utilities. (Bentham saw that this was mainly proper for "small quantities of pleasure".) Pigou distinguishes between the satisfaction that goods bring and their desirability. This became an important distinction to economists.[6]

Satisfaction is highly subjective. A main goal of utilitarianism was to maximize it for a population. Because it was very difficult to measure—many were doubtful that it could ever be scientifically measured—the economic community after 1920 sought to avoided it if possible. Desirability on the other hand was believed to be the driving force of choice behavior. It was objective and therefore more susceptible to scientific measurement and analysis. Following Marshall, Pigou in the above quote suggests that there are (i) a theoretical ratio relationship between the measurements of desire and satisfaction, and (ii) a testable ratio relationship between the measurements of desire and money.

[6] Current psychology (e.g., Kahneman, et al., 1997) makes the same distinction but with different terms. *Experienced utility* is an individual's direct estimates of her utility summed up as she experiences a hedonic episode. It is the utility resulting from Edgeworth's hedonimeter. It is distinguished from *decision utility,* which is the subject's estimates of utility for past hedonic episodes or for future ones.

Then ratios of satisfaction can be inferred if (i) is assumed and desire passes the test in (ii). This strategy makes utilitarianism dependent on the psychology relating desire and satisfaction, something that modern economics would like to avoid.

A scientifically better strategy for (ii) uses Section 7's indirect approaches to direct measurement. Utilitarians until relatively recently did not have access to these. Consider the following example which uses bisection measurement; similar examples can be given for magnitude production and estimation measurement. The subject is presented two hedonic items, a and c that he already has established a preference for, say $a \prec c$, and asked to "find a b that is midway between them in terms of desirability" (alternatively, "find a b such that the ratio of a to b is the same as the ratio of b to c"; or alternatively "find a b such that the difference from c to b is the same as the difference from b to a".) Using judgments of this kind, the axioms for bisection given in Section 7 can be checked for, and if they hold, then the data can be isomorphically modeled by a ratio scale, or equivalently isomorphically modeled by a translation scale having representations of the form $\varphi + r$ for each real r and any representation φ in the scale.

Of course, for money to be a good proxy for social utility, some linkage to individuals' preferences for pleasure or desirability is needed for its justification. The Utilitarians Bentham, Marshall, and Pigou failed to provide a credible one. Edgeworth, on the other hand developed the concept of the "core" of a qualitative exchange economy based on individual preferences for goods and conjectured that it would produce a quantitative perfect competitive market based on buying and selling through use of money. In the 1960s, various formulations of this conjecture was shown to be valid. However, even a perfect market has problems for intercomparing the utilities (as money) for individuals A and B, because it does it depends also on the utilities of other individuals in the society. Thus, if A and B were in a different population with a different perfect market, the order of their socially derived utilities with respect to that market could invert with respect to the previous perfect market.

11 Interpersonal Comparisons

For Wundt has shown that sensuous pleasures may thereby be measured, and, as utilitarians hold, all pleasures are commensurable. The first principle of this method might be: Just-perceivable increments of pleasure, of all pleasures for all persons, are equateable (cf. Wundt, Phys. Psych7., p. 295)

Edgeworth (1879), p. 369.

It is noteworthy that Jevons, Edgeworth, Robbins, and current economic theory agree that there is no empirical method to equate one person's pleasure with another's, but come to different conclusions about the reasons for this. Jevons considers such comparisons impossible because of the nature of the human mind. Robbins consider them outside of science. Current economic theory, which measures individuals' utilities through gambles and the utility theory developed by von Neumann-Morgenstern, consider them meaningless, because the theory has no way of establishing a unit to individuals' differences in utility. Edgeworth, however, believes they exist to a good approximation, but only in theory. We agree that empirically based interpersonal comparisons are impossible in most empirically based situations. But this does not mean that they are absolutely impossible, as suggested above. Later, an example is provided.

Let's start with Edgeworth. He took the equateability of jnds of pleasures across people as "a first principle incapable of proof." For over a century and a half, psychologists have grappled with the related issue of the equality of jnds for different senses within the same individual. In a recent review of psychophysical research Teghtsoonian (2012) writes, "What is being assumed is that a JND defined on one of those [physical] continua will be matched by a JND on the other. Indeed, it is an empirical question that, to my knowledge, has yet to be tested." So current psychological research is seemingly consistent with Edgeworth's position, *if we assume that pleasures of two individuals arise from separate physical continua.* If, however, we assume they arise from the same physical continuum, e.g., the pleasure obtained by consumption of a delicious ice cream, then Edgeworth is wrong about the non-empirical nature of the phenomenon. It will be argued that this extends to separate continua.

Central to Edgeworth's philosophy of Utilitarianism is the idea that individuals vary in their capacities for experiencing pleasure. Edgeworth (1879) writes,

> An individual has greater *capacity for happiness* than another, when for the same amount whatsoever of means he obtains a greater amount of pleasure, *and also* for the same increment (to the same amount) whatsoever of means a greater increment of pleasure. (p. 395.)

Suppose individuals i and ii purchase identical ice creams for the same price, and i received a higher number of pleasure jnds than ii. Then (for this ice cream continuum) i has greater happiness capacity than ii. Narens & Mausfeld (1992) analyzed things differently than Edgeworth. Applied to the ice cream continuum, they would argue that the properties of capacity for happiness described in the above quotation by Edgeworth could be achieved with i and ii satisfying Weber's Law with Weber constants c_i for i and c_{ii} for ii, and $c_i < c_{ii}$ for i having a greater capacity for happiness than ii. In such a situation, the holding of Weber's Law with $c_i < c_{ii}$ would constitute an empirical failure of Edgeworth's theory of just noticeable differences, because it requires, as mentioned in Edgeworth's epigraph given at the beginning of this section, that "just-perceivable increments of pleasure, of all pleasures for all persons, are equateable." Although this failure is within a continuum, it implies failure across continua, because a just-perceivable increment of pleasure of one continuum cannot be simultaneously equateable with two inequateable just-perceivable increments. More specifically, an individual iii on another hedonic continuum with Weber's

Law holding with Weber constant c_{iii}, cannot have c_{iii} equal to both c_{ii} and c_i.[7]

The above idea for comparing Weber constants on a single continuum like a flavor of ice cream can extend to comparisons across continua if the different continua can be *appropriately* mapped onto a single continuum. The following is an example.

Suppose $\mathfrak{X}_i = \langle X, \precsim_i \rangle$ is i's continuous semiorder for lots of goods X. Let $\langle X, \leq_{i,\star}, T_i \rangle$ be the continuous threshold structure induced by $\mathfrak{X}_i$. Let $\mathfrak{X}_{ii} = \langle X, \precsim_{ii} \rangle$ and $\langle X, \leq_{ii,\star}, T_{ii} \rangle$ be be similarly defined for ii. As a first step of mapping $\mathfrak{X}_i$ and $\mathfrak{X}_{ii}$ onto a common continuum, we assume that $\leq_{i,\star} = \leq_{ii,\star}$. This is a highly restrictive but empirically testable assumption. Let $\leq_\star = \leq_{i,\star} = \leq_{ii,\star}$. Assume a market for the lots in X. For each a in X, let $\mathsf{P}(a)$ be the price of a represented as a positive real number in the monitory unit used by the market. Let

$$T_i^\star = \text{the image of } T_i \text{ under } \mathsf{P}$$

and similarly for $T_{ii}^\star$. Then P is an isomorphism of $\langle X, \leq_{i,\star}, T_i \rangle$ onto $\langle \mathbb{R}^+, \leq, T_i^\star \rangle$ as well as an isomorphism of $\langle X, \leq_{ii,\star}, T_{ii} \rangle$ onto $\langle \mathbb{R}^+, \leq, T_{i,i}^\star \rangle$. For $\langle \mathbb{R}^+, \leq, T_i^\star \rangle$ and $\langle \mathbb{R}^+, \leq, T_{ii}^\star \rangle$ to have simultaneous jnd representations on a common jnd scale, Narens (1994) showed that necessary and sufficient conditions for this is that $T_i^\star$ and $T_{ii}^\star$ commute, that is for each x in $\mathbb{R}^+$,

$$T_i^\star[T_{ii}^\star(x)] = T_{ii}^\star[T_i^\star(x)]\,. \tag{14}$$

This is also empirical testable. Suppose that the commutativity in Equation 14 holds. Then a strictly monotonic transformation φ from $\mathbb{R}^+$ onto $\mathbb{R}^+$ can be found that simultaneously represents both $\langle \mathbb{R}^+, \leq_\star, T_{i,}^\star \rangle$ and $\langle \mathbb{R}^+, \leq_\star, T_{i,i}^\star \rangle$ onto a jnd scale. On this jnd scale, there exists a positive real constant r_i such that $T_i^\star(x) = x + r_i$ for all positive x, and similarly a positive real constant r_{ii} for $T_i^\star$. Thus the function Ψ, defined by,

$$\text{for all } a \text{ in } X,\ \Psi(a) = \varphi[\mathsf{P}(a)]\,,$$

maps X into a jnd scale where a jnd for T_i is r_i and a jnd for T_{ii} is r_{ii}. When $r_i \neq r_{ii}$, which is now an empirical matter, Edgeworth's view about equateablity of jnds for different pleasures is false, but his concept of "capacity for happiness" still holds.

It is highly unlikely that the conditions assumed above to get these equateablity results hold generally. But the point of the exercise was to demonstrate that arguments showing the widely held view in economics that is impossible to measure interpersonal comparisons of utility is flawed. While we agree that such measurements depend on strong empirical assumptions, and if tested in

[7]It is also worthwhile to note that it is unreasonable to expect that a person's capacities for happiness for hedonic experiences remain constant over time, for example, the jnd for a hedonic experience for the individual i at age fifteen remaining the same as when she is eighty five. With this consideration, individual ii in the above argument is not needed in order to carryout the above argument involving the non-equateability of a pair of jnds between i and iii.

a rigorous scientific manner, they are likely to fail. But this is not sufficient reason to assert their empirical impossibility.

Thus it is our conclusion that Jevons, Edgeworth, Robbins, current economic theory, and many other prominent economists and philosophers are wrong about the empirical nature of interpersonal comparisons because of wrong views about the nature of measurement. This is a result of having a perhaps too restrained views about what constitutes "intercomparability of utility". We next investigate weaker concepts of "intecomparability" that may be useful for defining "group utility" in terms of "individual utility".

Direct methods for the scientific measurement of utility have been routinely rejected by utilitarians and economists. The methods of magnitude production and estimation introduced and extensively examined empirically by Stevens and colleagues in 1950s and 60s were seen as suffering from the same deficiencies as other earlier attempts. However, as shown previously in Section 7, magnitude estimation and production have indirect measurement foundations. Experimental psychophysical research using these foundations have confirmed the Stevens School's findings of robust power law representations for psychophysical relations and their underlying assumption of ratio scales for psychophysical dimensions and power functions for psychological intensity, but generally with powers different from those obtained through Stevens' measurement methods. It is an open empirical matter whether these methods—when cleaned up by using the the Commutativity Property and Theorem 7 instead of the Multiplicative Property—extend to utility. We believe that it is likely that they do.

Some empirical research suggest that individuals' utility functions for money are power laws (e.g., Galanter, 1962). If so, then it follows that one person's individual's utility function for money is a power function on another's. This suggests that direct measurement techniques for utility such as the Equal Ratio Law might be of use for defining some utility intensity relationships across individuals for a pleasure continuum as well as a meaningful ordinal group utility function that is defined in terms individual utility scales. For many welfare considerations, the numerical size of the group utility is not needed; one needs to only pick a pleasure lot that is a maximum with respect to the the group ordering of pleasure lots.

It is assumed that utility for subjects i and ii have been appropriately measured through indirect versions of bisection, magnitude production, or magnitude estimation, or some other indirect method that produces ratio scales $\mathcal{S}_i$ and $\mathcal{S}_{ii}$ for respectively i and ii on a pleasure continuum $\langle X, \precsim_\star \rangle$. Let u_i and u_{ii} be i and ii respective utility representations from respectively $\mathcal{S}_i$ and $\mathcal{S}_{ii}$. Then the Equal Ratios Law states that for all a, b, c, and d in X,

$$\frac{u_i(a)}{u_i(b)} = \frac{u_i(c)}{u_i(d)} \text{ iff } \frac{u_{ii}(a)}{u_{ii}(b)} = \frac{u_{ii}(c)}{u_{ii}(d)} \,. \tag{15}$$

Note that in this Law, i is only making comparisons involving her utilities, and similarly ii is only making comparisons involving his. The experimenter, nature, or whatever, is relating these comparisons as logically equivalent. As discussed in Section 6, the Equal Ratios Law is logically equivalent to a power law relationship between i and ii; that is, there is a function φ from the positive reals onto the positive reals such that for each a in X there exists positive reals r and s such that

$$\varphi[u_i(a)] = ru_{ii}(a)^s \, .$$

Stevens (1948) suggests that the geometric mean is the appropriate statistic to use for averaging across ratios scales. Thus for a group n individuals, $1, \ldots, n$ with ratio scales $\mathcal{S}_1, \ldots, \mathcal{S}_n$ on X with representations, respectively, $u_1, \ldots, u_n$, a ratio scale $\mathcal{S}$ with a representation γ on X given by the following: There exists a positive real such t that for all a in X,

$$\gamma(a) = t \cdot [u_1(a) \cdot u_2(a) \cdot \ \cdots \ \cdot u_n(a)]^{\frac{1}{n}} \, . \tag{16}$$

$\gamma(a)$ can be viewed as representing the utility of an "average individual" of the society $\{1, \ldots, n\}$. Note that if the $u_1, \ldots, u_n$ in Equation 16 are the power functions,

$$u_1 = u_1, \ u_2 = r_2 u_1^{s_1}, \ \ldots \ , u_n = r_n u_1^{s_n} \, ,$$

then

$$\gamma = t \cdot u_1^{\frac{s_1 + \ \cdots \ + s_n}{n}} \, .$$

Also note that the ordering $\precsim$ on X given by the following, where γ is as in Equation 16, and for all a and b in X,

$$a \precsim b \ \text{ iff } \ \gamma(a) \leq \gamma(b) \, ,$$

is invariant under changes of representations from $\mathcal{S}_1, \ldots, \mathcal{S}_n$; that is, $\precsim$ is "meaningful" in the measurement-theoretic use of the word. Thus $\precsim$ is a meaningful interpersonal ordering for the population $\{1, \ldots, n\}$ that is produced without any direct comparisons of utilities of one individual with another. It can be used in welfare economics for finding a best option in X for the population $\{1, \ldots, n\}$.

Utilitarianism used the different ordering, $\precsim'$, for this purpose, where for all a and b in X,

$$a \precsim' b \text{ iff } u_1(a) + \ \cdots \ + u_n(a) \leq u_1(b) + \ \cdots \ + u_n(b) \, . \tag{17}$$

As indicated previously, Utilitarians considered various approaches for individual utility measurement. Following ideas of Bentham, Edgeworth considered both the hedonimeter and jnd approaches. Because of the existence of zero durations, the hedonimeter gave rise to representations forming a ratio scale. The jnd representation gives a representation fromt an absolute scale, whose single representation measures utility of a pleasure lot as the maximum number of happiness jnds produced by the lot.

Bentham and Edgeworth used Equation 17 as a means for maximizing individual and group utility. For Edgeworth's jnd approach this made sense, because of his assumption that "Just-perceivable increments of pleasure, of all pleasures for all persons, are equateable". This assumption allowed the number of jnds of pleasure that a population could experience given the resources to be the maximum group utility—the ideal goal of utilitarianism. But as argued previously, this way of reaching the goal is likely impossible, because jnds of pleasure across people need not be equateable. Bentham's and Edgeworth's other approach based on integration to individual measurement renders Equation 17 meaningless under changes of representations of individual scales. Assuming such changes of representation are forbidden theoretically or empirically, is tantamount to interpersonal comparisons of utility—something Bentham, Edgeworth, and the Utilitarians were unable to provide cogent arguments for, and which critics of Utilitarianism considered impossible. This issue is discussed more fully in Section 12.5.

12 Discussion

12.1 Utilities for pleasure and consumption

Utilitarians wanted to bring about policies to maximize the pleasure experienced by a population. To achieve this they needed methods to measure the satisfaction individuals experienced when they consumed available goods. Some Utilitarians, e.g., Marshal and Pigou, saw that it was easier to measure desire, because it was visible in economic activities such as buying and selling, and theoretically it could be linked to satisfaction by the previous quoted rule of Pigou (1920): "The ratio between the intensities of desire that he feels for the two is equal to the ratio between the amounts of satisfaction which their possession will yield to him". The problem with this rule is that it is often not applicable.

Because of the lack of rigorous methods for measuring satisfaction and increasing positivist tendencies toward science, other economic theorists around 1920 moved away from satisfaction as a key ingredient of economic theory. Instead, they sought to found their theory on the choices made by economic agents. These were linked to desirability by interpreting "b chosen over a" as the binary relation "a is less desirable than b", in symbols, $a \prec b$". The premise behind this was that desirability is the motivator for an economic choice, and such choices were observable. Thus, for describing the economic behavior of an individual, only desirability was needed, and therefore it was reasoned that utility should be a measure of the strength of intensity of desirability preference. This could be done through ordinal comparisons of desirability differences (Pareto, 1927).

In the 1960s, modern measurement theory developed axiomatic approaches to difference measurements. They were based on an observable quaternary ordering $\precsim$ on X, where $ab \precsim cd$ is interpreted as "the difference in preference between a and b is less than or equivalent to the difference between c and d. The following axiomatization and theorem is a special case of Theorem 2 of Chapter 4 of Krantz, et al. (1971) applied to a continuum.

Definition 2 $\langle X \times X, \precsim \rangle$ is said to be a *continuous difference structure* if and only if the following five axioms hold for all a, b, c, d, a', b', and c', in X, where ab, etc., are interpreted as the ordered pair (a, b), etc.

1. $\precsim$ is a weak ordering.

2. If $ab \precsim cd$, then $dc \precsim ba$.

3. If $ab \precsim a'b'$ and $bc \precsim b'c'$, then $ac \precsim a'c'$.

4. If $ab \precsim cd \precsim aa$, then there exists d' and d'' in X such that $ad' \sim cd \sim d'b$.

5. $\langle X \times X, \precsim \rangle$ is a weakly ordered continuum.

Theorem 8 *Suppose $\langle X \times X, \precsim \rangle$ is a continuous difference structure. Then there exists a real valued function φ on X such that for all a, b, c, and d in X,*

$$ab \precsim cd \text{ iff } \varphi(a) - \varphi(b) \leq \varphi(c) - \varphi(d)\,. \tag{18}$$

Furthermore, φ is unique up to a positive linear transformation, that is, if φ' satisfies Equation 18 in place of φ, then then there is a positive real r and real s such that $\varphi' = r\varphi + s$.

Although this system based on $\precsim$ applies to both the measurement of desire and satisfaction, it is more applicable in economics to the measurement of desire, because (*i*) desire is the driver of economic behavior and orderings of differences of desire derivable through observable purchasing behavior; and (*ii*) consumption is the basis of pleasure, and the orderings of the amounts of differences of pleasure derived from consumption is difficult to observe and get data about.

It should also be noted that for desire, $\precsim$ is often employed for ordering *commodity bundles*; whereas, for consumption it is usually employed as an ordering on *pleasurable episodes* that constitute the comsumptions of commodity bundles. Pleasurable episodes have a different internal structure and qualities than commodity bundles. In particular, they involve the experiencing of varying intensities of pleasure over a period of time. (The structure of episodes have been discussed in detail in Subsections 9.2.1 and 9.2.2.)

12.2 Utilities derived through risk

Utility was measured in terms of risk by Ramsey (1931). He considered gambles of the form $(a, \frac{1}{2}, b)$, "receiving a with probability $\frac{1}{2}$ or receiving b with probability $\frac{1}{2}$". Let X be the smallest set such that the *pure outcomes* $a, b, c, \ldots$ are in X, and if x and y is in X, then the *gamble* $x \otimes y = (x, \frac{1}{2}, y)$ is in X. Thus X consists of pure outcomes, $\frac{1}{2}$ gambles involving pure outcomes, or $\frac{1}{2}$ gambles, including $\frac{1}{2}$ gambles of $\frac{1}{2}$ gambles, etc. Ramsey presented axioms showing there exists an interval scale representation u such that for all a, b, c, and d in X,

$$\left(a, \frac{1}{2}, b\right) \precsim \left(c, \frac{1}{2}, b\right) \text{ iff } \frac{1}{2}u(a) + \frac{1}{2}u(b) \leq \frac{1}{2}u(c) + \frac{1}{2}u(d)\,. \tag{19}$$

Using modern measurement techniques, the assumption of a continuous bisection structure also yields an interval scale representation u satisfying Equation 19. Ramsey's result historically preceded the similar, highly influential one of von Neumann & Morgenstern (1947), but itself had no influence in economic utility theory until after von Neumann & Morgenstern.

The utility function u in Equation 19 derives its interval scale properties over pure outcomes through assumptions relating pure outcomes to gambles, particularly the consequence that the utility of receiving a pure outcome with probability $\frac{1}{2}$ is $\frac{1}{2}$ of the utility of the pure outcome. This is an assumption that Bentham explicitly makes. But Bentham does not employ it in establishing a utility function over pure outcomes: In Bentham's setup, the establishment depends only pleasure intensity and the amount of time spanning the pleasurable episode. In other words, risk is not a factor in establishment of Bentham-like utility but is in Ramsey and von Neumann & Morgenstern-like utility theories. Rationality considerations demand that for $\langle X, \precsim, \otimes \rangle$, utilities based on difference measurement of pure outcomes and Ramsey and von Neumann & Morgenstern-like utilities applied to the pure outcomes must coincide.

There has been criticisms of the von Neumann-Morgenstern as a rational system for measuring utility. The most prominent of these are Allais (1953) and Ellesberg (1961). They assert that risk hasn't been appropriately accounted for von Neumann-Morgenstern setup. But these and other objections have had limited impact on most economic thinking.

12.3 Utilities based on magnitude production and estimation

Direct measurement methods have always been suspect in Utilitarianism and economics. Stevens' theories of magnitude production and estimation were generally not accepted by later and current economists, because of lack of scientific rigor. But as discussed in Section 7, Stevens' methods can be modified to produce a useful and rigorous indirect counterpart. This counterpart can be used to measure individual utility. But like in the previous discussion concerning satisfaction, similar criticisms about it being *non-actionable* in the sense that magnitude production and estimation judgments do not occur in the making of choices about naturally occurring economic options. While this may be a valid criticism for decision analysis, a discipline that aims to uncover utilities for use in practical settings, it appears to us to be out of place for a normative theory utility based on rationality. The indirect counterpart of magnitude production can be used to produce a preference ordering over utility differences. Rationality would then demand that this ordering satisfies the previously discussed difference axioms. This appears to us to be a reasonable rationality assumption. Whether rationality holds empirically, or whether either magnitude production or difference measurement for utility hold empirically, are different matters.

12.4 Implications for Torgerson's Conjecture

Torgerson's Conjecture hasn't receive much attention in utility theory and philosophy. Part of the reason is that results about it are confusing. Some experiments show that subjects distinguish between ratios and differences, while many careful studies show they don't, but with the caveat that a minority of subjects usually do. Nevertheless, the experimental literature at least suggests that care should be taken in the interpretation of psychological experimental results involving ratios or differences as well as in ethical discussions involving thought experiments that use them.

There has been little theory as to why judgments of ratios and differences can be simultaneously modeled as ratios, or equivalently simultaneously modeled as differences. Torgerson's explanation that "The subject perceives only a single quantitative relation between stimuli," is merely a restatement in cognitive terms of the presented empirical findings. Narens (2000) provided the following measurement-theoretical explanation. It is based on the following theorem.

Theorem 9 *Suppose* $\langle X, \precsim \rangle$ *is a weakly ordered continuum and* F *is a strictly increasing function from* X *onto* X*. Suppose* φ *is a ratio scale representation that represents* X *onto* $\mathbb{R}^+$*,* $\precsim$ *onto* $\leq$*, and* F *onto* $F^\star$*. Suppose that* F *is meaningful with respect to* φ*'s scale, that is, for each positive* s*,*

$$F(s\varphi(x)) = sy = sF(\varphi(x))\,.$$

Then there exists r *in* $\mathbb{R}^+$ *such that for all* x *and* y *in* X*,*

$$F(x) = y \;\; \textit{iff} \;\; r = \frac{\varphi(y)}{\varphi(x)}\,.$$

Proof. Let s and t be arbitrary positive reals. Because φ is onto $\mathbb{R}^+$, let z in X be such that $t = \varphi(z)$. Then it follows from $\mathcal{S}$ being a ratio scale that for all s in $\mathbb{R}^+$,

$$F^\star(st) = F^\star([s\varphi(z)] = sF^\star[\varphi(z)] = sF^\star(t)\,. \tag{20}$$

Because Equation 20 is valid for all s and t, $F^\star$ is the functional equation on $\mathbb{R}^+$ satisfying $F^\star(st) = sF^\star(t)$. It is well-known that the only solution to this is $F^\star(t) = rt$ for some positive real r.

Psychologists generally assume or show that cognitive judgments are represented on a ratio scale. Theorem 9 shows that when such judgments form a strictly increasing function $F^\star$ that is invariant under ratio scale transformations, i.e., is meaningful, then it must be interpreted as a ratio. Thus a meaningful function $y = D^\star_s(x)$ representing the idea on a ratio scale that "the difference between y and x is s" can only be interpreted in meaningful way as a ratio, and except for certain degenerated cases, like $y = 4$ and $x = 2$, and therefore not as a general difference. This is the conclusion of Torgerson's Conjecture. It differs from Torgerson in that it assumes an underlying ratio scale and meaningfulness to analytically derive its conclusion; Torgerson assume an underlying ratio scale and uses empirical findings to support his conclusion.

In the proof of Theorem 9 it is critical that F is a function of a single variable. The generalization of the theorem to a function of two variables is not valid. The data Torgerson used in justifying his conclusion of "a single quantitive relation" usually involved functions of two or more variables, for example bisection, i.e., choosing the mid point m between a and b, is a function of two variables $m = M(a, b)$. This function can be interpreted as a relationship between two values of a function of a single variable, i.e.,

$$f(a) = m \text{ iff } f(m) = b\,.$$

Plateau gives this interpretation to his midpoint judgment as the cognitive-perceptual interpretation of what his artists were doing. In this case, Theorem 9 applies. However, there are other cognitive-perceptual interpretations where it is not proper to divide M into a single function relating values of it variables; e.g., view the placing of $M(a, b)$ as a perceptually symmetric solution involving a and b. It is an empirical matter whether these two computations give different results. Thus Theorem 9 provides limited support for Torgerson's *data,* while providing an independent argument for his conclusion.

By an exponential transformation, differences becomes ratios. Thus in isolation there is no qualitative difference between them. They are isomorphic. However, when presented simultaneously in the same ordered structure, rationality requires them to satisfy a distributive law. This restricts representing the ordered structure by a ratio scale. (See Theorem 6 of Chapter 2 of Krantz, et al. for an example involving an ordered semi-ring where distributivity forces an absolute scale.) While college subjects understand correctly the difference between ratios and differences when presented in instructions, they are not apparently able to correctly apply this knowledge in making subjective judgments of intensity. Let's assume that this is a property of human cognition that has to do with the fact that in making judgments of subjective intensity, the verbal understanding of quantity is different from the analog understanding of quantity used in making the judgment. This is not a far-fetched assumption given the current understanding of how different parts of the human brain process different kinds of information. If both kinds of quantity understanding are expressed verbally, like for example in a rating task, without their different origins being accounted for, then rationally based deductions using these mixed quantities could produce false conclusions. In our opinion, philosophical arguments involving utilities that lack formal foundations that fail to include psychological modeling of the processes involved utility estimation are particularly prone to this kind of concern.

12.5 Comparison of the Utilitarian Product with the Utilitarian Sum

A major goal of Utilitarianism was to maximize society's overall pleasure. We formalize this as a preference ordering of pleasure $\precsim$ over the set X of pleasurable items that the members of the society have in common. The goal then is to pick out a element x in X that is a $\precsim$-maximum given appropriate resource constraints. For the purposes of this subsection, we will consider the situation where the society has two individuals, (i) and (ii). The presented ideas and results generalize to $n \geq 2$ individuals. We also assume that individuals (i) and

(ii) have utility functions that take on positive values, u_i from a ratio scale $\mathcal{S}_i$, and u_{ii} from a ratio scale $\mathcal{S}_{ii}$. This covers many important distribution cases such as Utilitarian splitting of positive amounts of money.

As discussed earlier, Utilitarians thought that the population's preference ordering $\precsim$ should result from *Sum Utility;* that is, it results by the following definition: For all x and y in X,

$$x \precsim y \text{ iff } u_i(x) + u_{ii}(x) \leq u_i(y) + u_{ii}(y)\,. \tag{21}$$

For $\precsim$ to be a weak ordering, some form of intercomparability of utility between (i) and (ii) is needed for Equation 21 to be a valid definition for $\precsim$. Otherwise, different representations ru_i for various real r could be selected for i while leaving u_{ii} unchanged, which for fixed x and y and appropriate selections of r would invalidate Equation 21, making it *meaningless.* One way to avoid this is to allow linkages between $\mathcal{S}_i$ and $\mathcal{S}_{ii}$. Bentham attempted this by having the happiness of others affect the happiness of an individual, and Edgeworth added another dimension to intensity and duration of a pleasurable event to include the impact on the event's pleasure on others. Defining $\precsim$ using *Product Utility*,

$$x \precsim y \text{ iff } u_i(x) \cdot u_{ii}(x) \leq u_i(y) \cdot u_{ii}(y)\,, \tag{22}$$

does not have this kind of meaningfulness issue for positive utilities and does not need additional assumptions in order for the resulting social ordering $\precsim$ to be a well-defined weak ordering.

Three forms of distribution of society's resources has been extensively discussed in the literature:

- *Sum Utilitarianism* that maximizes the overall happiness of society by taking the sum or arithmetic mean of individual happinesses. (Both the sum and arithmetic mean produce the same ordering of items according to societal happiness.)
- *Prioritarianism* similar to Sum Utilitarianism except that worst off individuals are given priority in the distribution; and
- *Egalitarianism* that provides the same amount of happiness to all individuals.

A fourth form that has not been much studied in literature is

- *Product Utilitarianism* that maximizes the overall happiness of society by taking the product or geometric mean of individual happinesses. (Both the product and geometric mean produce the same ordering of items according to societal happiness.)

Product Utilitarianism has been suggested by Adler (2011) and others for its Prioritarian flavor, as is evident in the following examples of Skyrms & Narens (2018) that also show some of its advantages over Sum Utilitarianism and Egalitarianism.

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 maximization solution is $\frac{2}{3}\$100$ to Adam and $\frac{1}{3}\$100$ to Eve. Adam receives more money than Eve because Eve has a faster diminishing marginal utility for money than Adam.

In the last example, the Utilitarian Sum gives all the money to Adam because the assumed intercomparability between Adam and Eve utilities gives Adam greater utility for \$100 than it gives to Eve. Although this maximizes social utility, it appears unfair. Product Utility, which does not assume intercomparability, trades-off fairness with social utility and gives Adam a greater proportion of the \$100 than Eve, because Adam values \$100 more than Eve. This yields a higher socially utility than the more fair even split, but still gives Eve a reasonable share of the \$100.

If intercomparability fails, then Sum Utilitarianism fails rationality. Therefore, it should not be used as the normative Utilitarian rule for deciding social utility. If the right form of intercomparability holds, then the Utilitarian Sum is the obvious normative choice for carrying out the classical utilitarian program from Bentham through Edgeworth. But because of the discussed advantages of the Product Utility for defining social preference, it should be a serious candidate for normative utility aggregation in an Utilitarian setting. Because it incorporates a Prioritarian component into its maximization calculation, it may be helpful for mitigating some ethical concerns about Utilitarianism brought about by Sum Utilitarianism.[8] But Prioritarianism, unlike Product Utilitarianism, needs intercomparability to identify the "worst off". This is discussed in Subsection 12.8.

12.6 Subjective zeros in jnd measurement

As mentioned previously, Bentham had two utility theories: One based on integration of pleasure intensity over physical time, and one based on just-noticeable differences. The one based on integration had an objective 0 utility corresponding to integrals of episodes of 0 duration. These can be meaningfully equated across different individuals. What about hedonic jnd measurement? Are its episodes with neutral hedonic pleasure-pain also a meaningful utility 0 that is universal across kinds of episodes? More specifically, Should jnd measured hedonic episodes for a single individual that are judged neither positive (pleasurable) or negative (painful) have a meaningful 0 utility in theories where appropriate

[8] We know of only one experiment in which the Sum and Product Rules are compared for splitting of a windfall. It is in an experiment that compares peoples beliefs about justice in the splitting a resource similar to previous example involving Adam and Eve. In the article "On Dividing Justly", Yaari & Bar-Hillel (1984) presented the following question to 112 subjects.

"Q4: A shipment containing 12 grapefruit and 12 avocados is to be distributed between Jones and Smith. The following information is given, and is known also to the two recipients:

(1) Jones likes grapefruit very much, and is willing to buy any number of them, provided that the price does not exceed $1.00 per pound. He detests avocados, so he never buys them.

(2) Smith likes grapefruit and avocados equally well, and is willing to buy both grapefruit and avocado in any a number, provided that the price does not exceed $0.50 per pound.

(3) Jones and Smith are in the same income-tax bracket.

(4) No trades can be made after the division takes place. How should the fruits be divided between Jones and Smith, if the division is to be just?"

The results are in the following table, where "J:9–0" stand for Jones receiving 9 grapefruits and 0 avocados and "S:3–12" for Smith receiving 3 grapefruits and 12 avocados, etc.

Distribution			% of respondents
(1)	J:6–6	S:6–6	9
(2)	J:6–0	S:6–12	4
(3)	J:8–0	S:4–12	28
(4)	J:9–0	S:3–12	24
(5)	J:12–0	S:0–12	35

From a Utilitarian standpoint, Smith obviously should receive 12 avocados. This eliminates (1) as a proper utilitarian choice. Because Smith is going to receive 12 avocados as part of the utilitarian choice, such a choice then must then depend only on the split of grapefruits. The majority of subjects choose (5), the split of grapefruits obtained by maximizing the Utilitarian Sum. (4) is the split of grapefruits obtained by maximizing the Utilitarian Product. It is a sizable minority. The other sizable minority, (3), is close to it. Together they make a majority. Both (3) and (4) trade-off fairness with social utility in similar ways. Because of this, their closeness, and that together they are a majority, we conjecture that if all the subjects were force to choose between (4) and (5), then a majority would choose (4).

primitive predicates for "positive", "negative", and "neutral" hedonic experiences have been introduced?

Bentham's original intuition about hedonic jnd measurement was simple: Pleasures were a sum of atomic pleasures. Atomic pleasures are, by definition, the faintest perceivable pleasures. The atoms could be of different types, and the utility of a hedonic episode was the sum of the utilities of the atoms making up the episode's hedonism. There are two natural theories for determining the utility of atoms.

The first is the one expounded by Edgeworth in Section 11, where all jnds across all kinds of pleasures and all individuals are given the same value. We provided arguments against this and the principles upon which it was founded.

The second depends on the holding of the appropriateness of the continuous threshold measurement assumptions (Section 4) with the additional assumption that the items are perceive as having positive, neutral, of negative hedonic valences, with, in terms of each subject's induced (hedonic) weak preference ordering, that positive items are preferred to neutral ones which are preferred to negative ones. Then a jnd representation exists and there is a pleasure α with smallest judged positive jnd. By convention, α has utility value 1 on a jnd representation called called the *utile* representation. Then any other pleasure can be approximated within 1 jnd on that representation by multiplying the utility of α by an integer.

All neutral items will have values between 1, the value of α, and the largest value attained by a negative item ω on the utile scale. These neutral-valued items are perceived subjectively as indistinguishable in terms of amounts of total hedonism, and thus they are not be able to be ordered *phenomenologically* with respect to those amounts when compared against one another. But, because we are in a semiordered situation, they each can be compared with other items, and the *experimenter* can use such comparisons to order them. In short, although not phenomenologically distinguishable by the person, they may be behaviorally distinguishable by the person's responses to them *and other items.* This makes assigning utility 0 to all neutral items an unsound approach to jnd measurement, which is concerned with much more than representing direct comparisons of hedonistic experience. Items assigned utility 0 on the jnd utile scale have the unique property that all subjectively indistinguishable items are within 1 jnd from them. Although this property is meaningful for jnd measurement, it does not provide items assigned utility 0 a privileged phenomenological experience with respect to lack of positive or negative hedonic valence.

12.7 Product Utility with von Neumann-Morgenstern and Pareto

Results of Aczél and Roberts (1989) show that all meaningful forms of aggregation of individuals' ratio scales produce the same group ordering as the Utilitarian Product. This mathematical result raises rationality difficulties with Product Utilitarianism when its utility measurement incoporates probability.

So far, we have not used rationality assumptions regarding individual or group utility functions. Standard ones are:

- the individual and group utility functions satisfy the von-Neumann Morgenstern utility assumptions, and
- the group utility function is Weak Pareto.[9]

The following exam from Skyrms & Narens (2018) show Product Utility together with von Neumann-Morgenstern and Weak Pareto are incoherent. This is done by providing an example satisfying those assumptions whose gambles form a Dutch Book:

> Suppose that a Utilitarian's choices do not determine a forthcoming episode, but rather a gamble over possible episodes. How should she value such gambles? Bentham says to take the expectation. Suppose an individual's utilities are extended to probability distributions over possible episodes in this way.
>
> A social planner may also face choices of gambles over episodes. How should the Utilitarian planner value such gambles? Two possible approaches present themselves:
>
> (1) First, aggregate utilities of episodes by product; second take expectations to get utilities of lotteries.
>
> (2) First, individuals take expectations to get utilities of lotteries; second planner aggregates utilities of lotteries by product.
>
> These two approaches do not give the same result.
>
> We illustrate with a simple example:
>
> - Adam has Utilities 1, 3, 5 for prospects A, B, C respectively.
> - Eve has Utilities 5, 3, 1 from prospects A, B, C respectively.
> - There is also a lottery $< \frac{1}{2}A, \frac{1}{2}C >$, giving A or C, each with probability $\frac{1}{2}$.
>
> (1) first aggregates prospects by product, giving 5, 9, 5 to A, B, C. Then extends to lotteries by expectation, giving:

A	B	C	$< \frac{1}{2}A, \frac{1}{2}C >$
5	9	5	5

> (2) first has each individual extend to lotteries by expectation:

	A	B	C	$< \frac{1}{2}A, \frac{1}{2}C >$
Adam	1	3	5	3
Eve	5	3	1	3

> Then aggregates by product:

[9] Weak Pareto is an allocation state from which it is impossible to reallocate so as to make any one individual better off without making at least one other individual worse off.

A	B	C	$< \frac{1}{2}A, \frac{1}{2}C >$
5	9	5	9

Something is lost with each alternative:

- With alternative (2), the social planner is incoherent. Both A and C have social utility 5, but a 50%/50% gamble between them has utility 9 by the Product Utilitarianism. This violates the von Neumann-Morgenstern independence axiom.
- With alternative (1), the social planner loses consensus. Each individual is indifferent between preferring B to the lottery $< 1/2A, 1/2C >$. But the social planner strictly prefers B to the lottery. This violates a weak form of the Pareto principle.

The preceding shows that there are two, mutually exclusive, versions of Product Utilitarianism: The choice between (1) and (2) is a choice between group rationality (meaningfulness) and the respect for group consensus (Pareto). Skyrms and Narens conclude,

> Suppose one is not willing to give up either coherence or Pareto. And suppose one follows Bentham in extending utility to gambles by taking the expectation. Then a famous theorem of Harsanyi (1955) shows that one must be some version of a Sum Utilitarian. One then would have to wrestle with the meaningfulness problem that Product Utilitarianism solves.

12.8 Interpopulation Comparisons of Utility

It has been shown that Sum Utilitarianism fails for comparison of happiness in a population of ratio or interval scaled individuals without intercomparing utilities, but happiness comparisons can be achieved with ratio scaled individuals without intercomparisons through Product Utilitarianism. But what about comparisons of happiness between populations. The Prioritarian Parfit raised this difficulty for Sum Utilitarianism. He assumes utilities are real numbers, that that the boundary between “lives worth living” and “lives not with living” is 0 and that “lives barely worth living” identifies some positive real number. Using the Archimedean property of the reals, it then follows that the addition of a sufficiently large number of lives worth living will swamp in utility the sum of the utilities of individuals from any fixed population.

> 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 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 Benthamite account, individual utilities are measured not on an absolute scale but on a ratio scale. As we have seen, for a fixed population having positive utilities the Utilitarian Sum and arithmetic average are not meaningful, but the Utilitarian Product and geometric mean are. What about comparisons of utility across populations? Do measurement considerations still support Parfit's arguments?

First consider Product Utilitarianism. With Product Utilitarianism, adding lives (considered as episodes) whose utility is greater than 1 increases aggregate utility, while adding lives whose utility is between between 0 and 1, *decreases* the 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 neo-Benthamite treatment, zero utility is defined in terms of the null episode or equivalently, an instantaneous one. Parfit gets a zero utility as the division between lives worth living from those that are not. It also considers lives that are "barely worth living". It is this "barely" that causes meaningfulness problems in our neo-Benthamite approach. Parfit's fixing of 0 isn't a problem. The positive elements still form a ratio scale. But a range of values corresponding to "barely positive" does, because instances of this fixes the least upper bound of "barely", turning the ratio scale into an absolute one. Thus in our development of neo-Bentham utility theory, the utility comparisons used in Parfit's examples are meaningless. Nevertheless, our development together with

Product and Geometric Mean Utilitarianism captures the spirit of Bentham's views about extent and is consistent with the practical goals of Utilitarianism.

12.9 Accommodation Dynamics

No one could 'deny' interpersonal comparisons in the sense that they deny that people make them. Therefore those economists who 'deny' them must think that when a person says 'A is happier than B' he is deluding himself in thinking that he is making a statement of fact. But why should he be deluding himself? Why should it not be a statement of fact? It is probable that what is behind the idea that it is not a statement of fact, that one is not describing something one experiences, when one says 'A is happier than B', is some vague metaphysical doubt about the existence of minds other than one's own. . . . It is a mistake to suppose that another man's mind consists solely of feelings or images which one cannot ever experience (that is, that one's mind is a logical construction of personal feelings and images which are, by definition, not open to inspection by anyone else).

Little, 1957, pp. 54–55.

Theories of social justice require more than a subjective preference structure. They require an interpersonal comparison of welfare. Arrow agrees that we must take that step if we want to make judgements regarding social justice. That is, there must be some comparative information which is meaningful interpersonally. . . . Thus, for example, it must, at a minimum, be interpersonally meaningful to say that it is better to be Jack under circumstances α *than Jill under* β*. Sen argues that this is precisely the sort of judgement that justice must be built on: the sort of empathy which allows us to put ourselves in each others' shoes, and make comparative judgements regarding welfare. But precisely what form these judgements can and should take is less clear.*

Frohlich and & Oppenheimer (1999) p 10.

We agree with Little that people make interpersonal comparisons all the time. Further, we believe they form a basis for cooperative behavior which lead to productive group behavior. However, contrary to Little, we nevertheless believe that the comparisons as a description of reality is a delusion. Our goal in this subsection is to describe how such a delusion arises and why people are likely to mistake it for reality. We follow the development of Narens & Skyrms (2017) that show how a simple dynamic of a couple adjusting their perspectives of how the other couple member values hedonic actions leads to an equilibrium where the couple agrees as to what action would produce the higher joint benefit. Narens & Skyrms view this agreement as establishing a convention similar to others that people and society establish for initiating an maintaining cooperative behavior. The convention established is just one of many possible. It depends not only on each person's utility for hedonic actions, but also on incidental matters such as the order in which joint actions are experienced, their dynamic choices of adjustment parameters, etc. Narens & Skyrms considers the couple's belief that this form of understanding each others values as a true comparison of values is a delusion.

They start out assuming a couple, Adam and Eve, have complete knowledge of their own utility preferences over hedonic actions each can do, for example, dish washing. At this point preferences are assumed to be egoistic and do not yet incorporate the other-regarding elements such as Adam's understanding of how Eve values dish washing. It is further assumed that they they evaluate their own utility preferences rationally through von Neumann-Morgenstern utility functions, u for Adam and v for Eve for these egoistic actions. These utility functions are from interval scales and thus each has an arbitrary 0 value and an arbitrary unit. Knowledge of the scales by themselves does not allow Adam and Eve to make the common sense interpersonal judgments. The 0's are not a problem, because Adam and Eve want to compare differences like: Is Adam's decrement in utility due to his washing small in comparison to his view of Eve's increment in utility due to not washing? In comparing differences, choice of the 0's wash out. Units are the problem. Different choice of units can make Eve's perceived difference arbitrarily greater or smaller than Adam's. They have perfect knowledge of each other's preferences, because we assume they are given by the other each other's preferences. This does not help: Although each knows the other's utility scale up to 0 and unit, they don't know how to make the trade-off between their utility differences. Without such knowledge they can't intercompare differences.

Narens & Skyrms posit that they extend their individual utility functions to incorporate a *view* of the other's utility function. It is important to note that they are not incorporating the other's utility function, only a view—a perception or model—of it. Unlike one's own and the other's utility functions, which are assume to be fixed throughout, this view can change with time; that is, it can be updated to incorporate new information. Adam's view u_t of Eve's utility function v at time t is related to v by the formula,

$$u_t = a_t v\,,$$

where a_t is a positive constant. This is reasonable, because by complete knowledge Adam knows v (because Eve has given him this information) and he knows that to put v into his units u_t must be some constant multiple of it. (He knows this because he is rational). Thus he begins at time 0 with an estimate $a_0 v$ as u_0 as his first estimate of v *in his units for* u. Eve does similarly for her view of Adam's utilities with

$$v_t = e_t u \text{ and } v(0) = e_0 u\,.$$

Narens & Skyrms assume that Adam and Eve take initial values a_0 and e_0 as,

$$a_0 = \frac{u_0(E_{\text{max}}) - u_0(E_{\text{min}})}{u(A_{\text{max}}) - u(A_{\text{min}})} \text{ and } e_0 = \frac{v_0(A_{\text{max}}) - v_0(A_{\text{min}})}{v(E_{\text{max}}) - v(E_{\text{min}})}\,. \tag{23}$$

Note that a_0 and e_0 are meaningful; that is, if instead of u in Equation 23 another representation $ru - s$, $r > 0$ and s is an arbitrary real, is taken from

Adam's scale family, then Equation 23 will still hold, and similarly for e_0 and Eve. The distance between Adam's and Eve's tradeoff constants at time 0 is $\delta_0 = |a_0 - e_0|$. By Equation 23 the difference (or distance) δ_0 is meaningful.

Times are assumed to be integral and nonnegative. Except for possibly the initial time 0, they mark disagreements, If at time t Adam and Eve disagree about tradeoffs, for example Adam believes action B has more utility for him then C has for Eve,

$$u(B) > u_t(C)\,,$$

but Eve believes otherewise,

$$v(C) > v_t(B)\,,$$

this is a disagreement at time t. When there is a disagreement it is assume they accommodate by changing their utility views of the other by moving their tradeoff constants closer together. If there is no disagreement, there is consensus as to what is the best action to jointly take. In case of a disagreement, Narens & Skyrms assume they change their views of the other's utility by just moving their view of the other's utility closer to their own utility function. Thus if there is any difference, they will, depending on the difference, move closer together to reduce the difference. Narens & Skyrms formulate this as follows:

> For the first disagreement Adam moves a_0 in the direction of e_0 (but not equaling or surpassing it) by choosing a positive real number p that does not depend on Adam's scale and making $a_1 = pa_0$ his new tradeoff constant. Because a_0 is [meaningful] and p does not depend on Adam's scale, a_1 is [meaningful]. Similarly Eve obtains an [meaningful] tradeoff constant $e_1 = qe_0$. Resolving other disagreements then lead to the sequence of [meaningful] tradeoff constants, a_i, e_i, as i ranges over the nonnegative integers. This gives rise to an accommodation dynamics that maps the difference $|a_i - e_i|$ into the smaller difference $|a_{i+1} - e_{i+1}|$ except when $|a_i - e_i| = 0$.

Narens & Skyrms then shows that this accommodation dynamics converges (Theorem 10):

Formally, call a function f from the positive real numbers into the nonnegative real numbers an *accommodation function* if and only if

(i) f is continuous,

(ii) $f(d) < d$ if $d > 0$, and

(iii) $f(0) = 0$.

An *accommodation dynamics* has an accommodation function f and starts from an initial positive real d_0 producing the *accommodation sequence,* $d_0, d_1 = f(d_0), d_2 = f(f(d_0)), \ldots$.

We assume that $\delta_i = |a_i - e_i|$ is an accommodation sequence with accommodation function g such that $g(\delta_i) = \delta_{i+1}$.

Theorem 10 *The sequence δ_i converges to an equilibrium in which $\delta = 0$.*

Theorem 10 shows that δ_i converges to an equilibrium in which there is agreement on interpersonal comparisons of utility, that is, converges to $\delta = 0$.

Narens & Skyrms conclude their article the following:

> Narens and Luce [1983] suggest that comparability of utilities is based on self-deception. Individuals evolve conventional trade-offs, and then mistake them for objective fact. After all, the propensity of humans to mistake their own conventions for objective reality is well-known to anthropologists. Accommodation dynamics can provide some insight on the evolution of such conventions.
>
> . . . Whether peoples' utility can be validly compared is a central issue in philosophy and economics. The usual argument against the intercomparisions assumes utilities come about through von-Neumann-Morgenstern expected utility theory and thus are represented on an interval scale. It is then argued that because there is no principled way of identifying multiplicative units and constants across different people's utility scales, valid intercomparisons are impossible (e.g., Robbins, 1935, 1938). The usual argument for intercomparibility is that examples of it are observed all the time (e.g., Little, 1957). We find both kinds of arguments are deficient: The "against" doesn't explain why the rationality or the psychology inherent in von Neumann-Morgenstern cannot be extended in ways resulting in valid interpersonal comparisons. The "for" argument fails to consider alternative explanations for observed agreement behaviors regarding utilities that do not rely on an underlying truth.
>
> Here we show that even given the best environment for interpersonal comparisons using extended utilities [i.e., views of other's utilities], a wide class of accommodation dynamics can account for observed behavior. They converge to an equilibrium at which individuals agree, but there are an infinite number of such equilibria and which one is reached depends on the starting point, details of the dynamics, and the occasioning the adjustments. The equilibria are invariant under change of scale. They provide a principled way of identifying scale units and constants throughout the accommodation process as they converge to an invariant interpersonal equilibrium.
>
> This equilibrium can be viewed as a social contract and such contracts can increase the overall utility for the individuals involved by encouraging collective action (e.g., see Skyrms, 1996). We believe that it is more productive to view interpersonal comparisons as social contracts instead of as a means for trying to achieve some social optimization or to carry out some moral imperative. Having many or conflicting equilibria are not an impediment for achieving an agreeable social contract.

13 Conclusions

We have seen that the hedonic Utilitarianism of Bentham is not the dead end that many have taken it to be. Theoretical possibilities exist for delivering many, if not all, of its goals. There are live open questions. Many turn on empirical tests that have not yet been carried out.

References

Aczél, J., & Roberts, F. S. (1989). On the possible merging functions. *Math. Soc. Sci., 17,* 205–243.

Adler, M. (2011). *Well-being and Fair Distribution: Beyond Cost-Benefit Analysis.* New York: Oxford University Press.

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.

Bailey, S. (1825). *A Critical Dissertation on Value.* London, p. 5.

Birnbaum, M. H. and Mellers, B. A. (1978). Measurement and the mental map. *Perception & Psychophysics, 23,* 403–408.

Bernasconia, M., Choiratb, C., Raffaello, S. (2008). Measurement by subjective estimation: Testing for separable representations. *Journal of Mathematical Psychology, 52,* 184–201.

Dzhafarov, E. N. & Colonius, H. (2011). The Fechnerian idea. *The American Journal of Psychology, June,* 1–18. DOI: 10.5406/amerjpsyc.124.2.0127.

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

Edgeworth, F. Y. (1881, 1887). *Mathematical Psychics.* London: G. Kegan Paul, Appendix III "On Hedonimetry".

Ellsberg, D. (1961). Risk, ambiguity and the Savage axioms. *Quarterly Journal of Economics, 75,* 643–649.

Ellermeier, W. and Faulhammer, G. (2000). Empirical evaluation of axioms fundamental to Stevens's ratio-scaling approach: I. Loudness production. *Perception & Psychophysics, 62(8),* 1505–1511.

Falmagne, J-C. (1985). *Elements of Psychophysical Theory.* Oxford: Oxford University Press.

Fechner, G. T. (1887). Über die psychischen Massprincipien und das Weber'sche Gesetz. *Philosophische Studien, 4,* 161–230.

Fechner, G. T. (1887). *Sachen der Psychophysik.* Leipzig. (p. 215).

Norman Frohlich, J. & Oppenheimer, J. (1999). Kenneth Arrow, Welfare Aggregation and Progress in Political Theory. In James Alt, J., Levi, M., & Elinor Ostrom, E, Eds, *Competition and Cooperation.* New York: Russell Sage Foundation, Pp. 4–33.

Galanter, E. (1962). The direct measurement of utility and subjective probability. *American Journal of Psychology, 75,* 208–220.

Garner, W. R. (1954). A technique and a scale for loudness measurement. *J. Acoust. Soc. Am., 26,* 73–88.

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

Krantz, D. H., Luce R. D., Suppes, P., and Tversky, A. (1971). *Foundations of Measurement, Vol. I.* New York: Academic Press.

Hicks, J. R. (1939). *Value and Capital: An inquiry into some fundamental principles of economic theory (2d edition).* Oxford: Clarendon.

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

Little, I. M. D. (1957). *A Critique of Welfare Economics, 2nd ed.* London: Oxford University Press.

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

Luce, R. D., Steingrimsson, R., & Narens, L. (2010) Are psychophysical scales of intensities the same or different when stimuli vary on other dimensions? Theory with experiments varying loudness and pitch. *Psychological Review*, 2010, 117, 1247–1258.

Mitchell, W. C. (1918). Bentham's Felicific Calculus. *Political Science Quarterly, Vol. 33, No. 2 (Jun., 1918),* 161–183. Stable URL: http://www.jstor.org/stable/2141580 Accessed: 27-10-2016 00:49 UTC

Narens, L. (1994). The measurement theory of dense threshold structures. *Journal of Mathematical Psychology, 38,* 301–321.

Narens, L. (1996). A theory of magnitude estimation. *Journal of Mathematical Psychology, 40,* 109–129.

Narens, L. & Mausfeld, R. (1992). On the relationship of the psychological and the physical in psychophysics. *Psychological Review,99,* 467–479.

Narens, L. & Luce, R. D. (1983). Why we may have been mislead into believing in the interpersonal comparability of utility. *Theory and Decision,* 247–260.

Narens, L. & Skyrms, B. (2017). *Philosophical Studies, https://doi.org/10.1007/s11098-017-0966-6.*

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

Pareto, V. (1927). *Manuel d'Economie Politique, 2nd ed.* Paris: Giard.

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

Pigou, A. C. (1920) The Economics of Welfare. Reprint, 2002, New Brunswick, N.J.: Transactions Press.

Plateau, M. J. (1872). Sur la measure des sensations physiques, et sur la loi qui lie l'intensité de ces sensations à l'intensité de la cause excitante. *Bulletin de l'Académie Royale des Sciences, es Lettres et des Beaux Arts de Belgique, 33,* 376–388.

Ramsey, F. P. (1926) Truth and Probability. In Ramsey, 1931, *The Foundations of Mathematics and other Logical Essays, Ch. VII,* p. 156–198, edited by R. B. Braithwaite. London: Kegan, Paul, Trench, Trubner & Co., New York: Harcourt, Brace and Company.

Robbins, L. (1932). *An Essay on the Nature and Significance of Economic Science.* London: MacMillan & CO.

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

Savage, L. J. The Foundations of Statistics, John Wiley and Sons, New York (1954).

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

Skyrms & Narens (2018). Measuring the Hedonimeter. *Manuscript.*

Skyrms & Narens (2018). Two kinds of product utilitarianism. *Manuscript.*

Steingrimsson R. & Luce R., D. (2012). Predictions from a model of global psychophysics about differences between perceptual and physical matches. *Attention, Perception, & Psychophysics. 74:1668?1680, DOI 10.3758/s13414-012-0334-8.*

Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103,* 677–680.

Stigler, G., J. (1950). The development of utility theory II. *Journal of Political Economy, 58, No. 5,* 373–396. (From Stigler II pages 380-381.)

Teghtsoonian, R. (2012). The standard model for perceived magnitude: A framework for (almost) everything known about it. *American Journal of Psychology, 125, No. 2,* 165–174.

Torgerson, W. S. (1961). Distances and ratios in psychological scaling. *Acta Psychologica, 19,* 201–205.

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

Walras in 1874 and 1877 published Éléments d'économie politique pure (1899, 4th ed.; 1926, éd. dfinitive), in English, Elements of Pure Economics (1954), trans. William Jaffé.

Yarri, M. E. & Bar-Hillel, M. (1984). On dividing justly. *Social Choice and Welfare, 1,* 1–24.

Zimmer, K. (2005). Examining the validity of numerical ratios in loudness fractionation. Perception & Psychophysics, 67, 569?579.