DRAFT

Is it Possible to Measure Happiness?

The measurement-theoretic argument against subjective measures of wellbeing

Erik Angner[†‡]

† I am most grateful to Zvi Biener, Greg Frost, Brian Hepburn, and Peter Machamer for constructive comments on earlier versions of this paper.

‡ To contact the author write to Dept. of Philosophy and Dept. of Economics, University of Alabama at Birmingham, 1530 3rd Avenue South, HB 414A, Birmingham, AL, 35294-1260; e-mail: angner@uab.edu.

Is it Possible to Measure Happiness?

The measurement-theoretic argument

against subjective measures of wellbeing

Word Count: 11412

This paper examines what may be the most common argument against subjective measures of wellbeing. The argument is based on the claim that happiness does not, whereas preference satisfaction does, permit the development of adequate measures. The aim of this paper is twofold: first, to identify the central assumptions on which the argument rests; second, to assess how convincing it is. The analysis suggests that the argument against subjective measures rests on an empirical premiss, viz. that people's choices in fact satisfy the axioms of rational choice theory. Because this claim is increasingly difficult to defend in light of recent developments in behavioral economics, I conclude that the argument as it stands is unconvincing. However, the failure of this argument does not imply that degrees of happiness are easier to measure than degrees of preference satisfaction. Indeed, I argue that the emphasis on the measurement-theoretic argument obscured several very real problems associated with subjective measures.

---

1. Introduction

The last couple of decades have witnessed the rise of so-called subjective measures of wellbeing, designed to reflect the degree to which an individual is happy, satisfied, or the like. Unlike the more widely used economic measures, which are typically based on data about income, market transactions and the like, subjective measures are based on answers to questions such as: "Taking things all together, how would you say things are these days – would you say you're *very happy*, *pretty happy*, or *not too happy* these days?" (Gurin, Veroff, and Feld 1960, 411, italics in original). The subjective measures are typically presented as alternatives to, and improvements over, traditional economic measures for purposes of the articulation and evaluation of public policy. Since the new measures suggest rather different answers to questions about the determinants and distribution of wellbeing or welfare,[1] a shift to subjective measures for public policy purposes could have vast implications.[2]

The new measures have also attracted a great deal of criticism, especially from economists who are comfortable using the more traditional measures. The aim of this paper is to assess what may be the most common argument against subjective measures of wellbeing. The argument is based on the notion that the degree to which people are happy or satisfied cannot be adequately measured. The argument appears e.g. in Wilfred Beckerman's book *Two Cheers for the Affluent Society: A spirited defense of economic growth* (1974), which defends the use of Gross National Product (GNP) as a measure of welfare. Beckerman recognizes that there are different views of what constitutes welfare, and he acknowledges that some people think welfare is a matter of happiness. Nevertheless,

---

[1] Following standard practice, I use these terms interchangeably (see e.g. Sumner 1996, 1).

[2] For useful overviews of subjective measures, see e.g. Diener et al. (1999), Kahneman, Diener, and Schwarz (1999) and Diener, Lucas, and Oishi (2002). For a more extended discussion from a philosophical perspective, see Angner (2005).

he insists on using GNP as a measure of welfare. Beckerman's main reason is that "[the] concept of happiness is one for which there can be no scientific objective measure" (Beckerman 1974, 53).

While it is clear that Beckerman hopes to demonstrate the superiority of traditional economic measures over measures of wellbeing-as-happiness (and the like), the exact shape of the argument remains unclear. Presumably, he recognizes that economists and psychologists operate with different accounts of wellbeing: whereas economists adopt some kind of preference satisfaction account, psychologists rely on some kind of mental state account. The idea, surely, is that the former permits the development of adequate measures of wellbeing, whereas the latter does not. Here, I will not quarrel with the inference from 'Mental states like happiness cannot be measured' to its corollary 'Subjective measures are inadequate.' Instead, I will focus on the argument to the effect that mental states like happiness do not, whereas preference satisfaction does, permit the development of adequate measures. For reasons that will be clear, I will refer to this as the *measurement-theoretic argument* against subjective measures.

Key to understanding the exact shape of this argument – and the disagreement between economists and psychologists more generally – is the recognition that they operate with radically different approaches to measurement. As David H. Krantz (1991) notes, there are broadly speaking two approaches to measurement in the behavioral and social sciences. In his words: "One, which may be termed the psychometric approach, introduces latent [unobservable] variables to explain behavioral orderings. The second ... treats the numerical representation of behavioral orderings axiomatically" (Krantz 1991, 2). Because the latter relies on the so-called theory of measurement, I will refer to the it as the *measurement-theoretic* approach to measurement. In sections 2 and 3, I will argue that while psychologists operate within the psychometric approach, economists operate within the measurement-theoretic one. This fact, I claim, sheds a great deal of light on psychologists' and economists' efforts to measure wellbeing, and in particular helps us identify the exact shape of the argument against subjective measures.

The main conclusion of this paper, articulated in section 5, is that the measurement-theoretic argument remains unconvincing. This is so because it relies critically on an empirical premiss – viz. that people's choices in fact satisfy the axioms of rational choice theory – which is increasingly difficult to maintain in light of recent advances in behavioral economics. The conclusion does not mean, however, that e.g. degrees of happiness in fact are easier to measure than degrees of preference satisfaction, or that subjective measures are superior to traditional economic measures for policy purposes. Indeed, I will argue, the measurement-theoretic argument obscures the truly troubling features of subjective measures. In the concluding section 6, I identify what I think those issues are.

The measurement-theoretic argument is interesting for several reasons. To begin with, it represents an obvious challenge for proponents of subjective measures. If the argument is sound and mental states do not permit the development of adequate measures, the whole "science of happiness" would suffer a devastating blow. That is, it would seem that subjective measures could not legitimately be used to identify the determinants and distribution of wellbeing. For the same reason, it would appear, they could not be used as a guide in the articulation and evaluation of public policy. Thus, the reorientation of public policy envisioned by psychologists working on subjective measures would not be justified.

Moreover, it is sometimes argued that accounts of wellbeing can be judged in part on the grounds of whether they permit the development of adequate measures of wellbeing. This idea is explicit in several prominent contemporary philosophers. James Griffin argues that we cannot "first fix on the best account of 'well-being' and independently ask about its measurement. One proper ground for choosing between conceptions of well-being would be that one lends itself to the deliberation that we must do and another does not" (Griffin 1986, 1). The idea, presumably, is that the measurement of wellbeing is necessary for the deliberation that we have to do. Similarly, Christine M. Korsgaard maintains that an account of the quality of life may be assessed "for its utility in determining actual political and economic policy – that is, whether it provides accurate enough measures to assess the effects of policy" (Korsgaard 1993, 54). If Korsgaard is correct, the question of what can be adequately measured is eminently relevant to the adequacy of accounts of wellbeing, and to the plausibility of the ethical theories in which that concept plays a role.

2. The psychologists' approach to measurement

In this section, I explore what makes psychologists think that they are able to measure mental states like happiness and satisfaction. My main proposition is that psychologists involved in efforts to measure wellbeing operate within the psychometric approach. The proposition, I claim, is supported by the fact that the science of happiness is best seen as a branch of personality psychology, and by the fact that personality psychology has always been intimately linked with psychometrics. The proposition also has a great deal of explanatory power. Among other things, it helps account for the psychologists' attitudes toward economic measures.

2.1 Personality psychology

In spite of the fact that subjective measures are often described as a recent development, their history goes back at least to the late 1920's and early 1930's.[3] During this period, a number of social workers, psychologists, and psychiatrists – including Katharine Bement Davis (1929) and G. V. Hamilton (1929) – simultaneously began to explore happiness systematically in the context of studies on "marital adjustment." Shortly afterwards, the idea was picked up by educational

[3] In this section, I follow the exposition of Angner (manuscript).

psychologists like Goodwin Watson, a Columbia professor of education. He motivates his project in the following way:

> No human quest may claim a larger following that that for happiness and satisfaction in life. Even the highest ethics tends to justify itself by its contribution to human happiness.... It becomes, therefore, extraordinary almost beyond belief that so few attempts have been made to apply the techniques of psychological study to the understanding of happiness (Watson 1930, 79).

By "the highest ethics," presumably, Watson refers to the utilitarians' maxim of the greatest happiness for the greatest number. Either way, there is little doubt that he himself is intensely interested in using "psychological techniques" for the purposes of promoting happiness.

The sudden appearance of the systematic study of happiness and satisfaction should be seen in light of the emergence of personality psychology – the study of individual differences, or "human individuality" (McAdams 2001, 11308) – as an independent subfield of psychology after World War I. In O. P. McAdams' words: "Building on the pioneering work of Francis Galton and Alfred Binet on mental testing and spurred by the mobilization of large military forces in World War I, psychologists began to invent self-report, multi-item tests to assess individual differences in personality functioning" (McAdams 2001, 11309). Driven in part by military and other applications, psychologists explored various methods to measure individual characteristics, and as a result of their efforts, the early personality psychologists ultimately convinced themselves that they were able to do so (McAdams 2001, 11309).

It is hardly a coincidence that efforts to measure happiness got off the ground as personality psychology emerged as an independent subfield of psychology. It would make little sense for psychologists to try to measure how happy people are before they have convinced themselves that they can measure a variety of personality characteristics. This insight allows us to address Watson's puzzle in the quote above, regarding why it had taken so long since the articulation of the utilitarian principle for psychologists to explore systematically who is happy and why. The Watson quote also suggests an answer to the question of why the effort to measure happiness got underway relatively shortly after the emergence of personality psychology. Assuming that many psychologists of the time had strong utilitarian tendencies, it is easy to see why some of them would want to measure happiness.

2.2 The psychometric approach

Personality psychology has always been intimately linked with psychometrics. As David G. Winter and Nicole B. Barenbaum noted, from the very beginning personality psychology was "devoting an extraordinary amount of attention to issues of measurement and psychometrics" (Winter and Barenbaum 1999, 5). These efforts were in large part driven by a desire to be useful to corporations, the government and the military (Winter and Barenbaum 1999, 5). What characterizes the psychometric approach, other than the fact that it "introduces latent [unobservable] variables to explain behavioral orderings" (Krantz 1991, 2)? In order to develop a better idea of the nature of the psychometric approach, it is reasonable to have a look at standard textbooks such as *Psychometric Theory* (Nunnally and Bernstein [1967] 1994) and *The New Psychometrics* (Kline 1998).

A central term of the psychometric approach is *reliability*. As Kline puts it, the notion of reliable measurement means that it is "without variation regardless of when the measurement is made or who makes the measurement, provided only that the individual [i.e. the person taking the measurement] is sane, in possession of his or her faculties and trained to use the instrument" (Kline 1998, 26). There are at least two kinds of reliability. A measure has *test-retest reliability* insofar as it yields "the same score for each subject when he or she takes the test on another occasion, given that their status on the variable has not changed" (Kline 1998, 29). A test has *internal consistency* insofar as "each item [of the test administered] measures the same variable" (Kline 1998, 30). A desire to prove the reliability of subjective measures, I would suggest, explains both the concern with stability over time (which is in essence the same thing as test-retest reliability) and internal consistency. If total test scores varied too much over time, or individual items of the test diverged too much from each other, there would be reason to think that the measures were unreliable.

Another central term in the psychometric approach is that of *validity*. As Kline puts it: "A test is said to be valid if it measures what it purports to measure" (Kline 1998, 34; cf. Nunnally and Bernstein 1994, 83). Presumably, the locution "what [the test] is supposed to measure" refers to the attribute that the author or user of the test (as the case may be) takes the test to measure. Again, there are several types of validity, including *face validity, concurrent validity, predictive validity,* and *content validity* (cf. Kline 1998, 34-37) Here I will focus on *construct validity,* which is the form of validity most relevant to the present purposes. This idea, which was introduced by Cronbach and Meehl (1955), is explained in more detail by Nunnally and Bernstein (1994).

Introducing the topic, Nunnally and Bernstein write: "All sciences, including psychology, are concerned with establishing functional relations between important variables" (Nunnally and Bernstein 1994, 84). They continue:

> To the extent that a variable is abstract and latent rather than concrete and observable ... it is called a "construct." ... A construct reflects a hypothesis (often incompletely formed) that a variety of behaviors will correlate with one another in studies of individual differences and/or will be similarly affected by experimental manipulations. Nearly all theories concern statements about constructs (Nunnally and Bernstein 1994, 85).

Furthermore, they claim: "In general, science's two major concerns are (1) developing measures of individual constructs and (2) finding functional relations between measures of different constructs" (Nunnally and Bernstein 1994, 85). The concept of construct validation, which comes into play when "measuring psychological attributes" (Nunnally and Bernstein 1994, 83), is intended to help scientists reach these goals. Construct validation, according to Nunnally and Bernstein (1994), has three steps. As the two authors put it:

> There are three major aspects of construct validation: (1) specifying the domain of observables related to the construct; (2) determining the extent to which observables tend to measure the same thing, several different things, or many different things from empirical research and statistical analyses; and (3) performing subsequent individual differences studies and/or experiments to determine the extent to which supposed measures of the construct are consistent with "best guesses" about the construct (Nunnally and Bernstein 1994, 86-87).

In what follows, I will discuss the three steps in order. There is no assumption, however, that the three steps take place in this order in the context of actual research.

First, scientists need to identify a class of observable variables that are related to the construct. Nunnally and Bernstein argue that there is no precise method that one can follow in this step. As a result, the scientists must therefore to a great extent rely on intuition and preconceived ideas about how the construct would vary across conditions. Having identified such a class is, nevertheless, a necessary condition for the second step, which is to explore whether the observable variables can be described as measuring the same thing or not. This step is performed by "determining how well the measures of observables 'go together' (intercorrelate) empirically" (Nunnally and Bernstein 1994, 88). This means that the scientist needs to collect data about how the observables vary across conditions, and compute the relevant correlation coefficients. The authors add:

> The results of investigations like those described above lead to one of three conclusions. If all the proposed measures correlate highly with one another, it can be concluded that they all measure much the same thing. If the measures tend to split up into clusters such that the members of a cluster correlate highly with one another and correlate much less with the members of other clusters, they measure a number of *different* things.... A third possibility is that the correlations among the measures all are near zero, so that they measure different things and there is no meaningful construct (Nunnally and Bernstein 1994, 90).

The third step is to show that a set of highly correlated observables in a domain can legitimately be taken to be measures of the construct in which the scientist is interested. In the words of Nunnally and Bernstein: "To the extent that the elements of such a domain [are intercorrelated], *some* construct may be employed to account for the interrelationships, but it is by no means certain that the construct name which motivated the research is appropriate" (Nunnally and Bernstein 1994, 90). To see whether a set of intercorrelated variables can be assumed to be a measure of a given construct (like anxiety, stress or happiness) the scientist needs to explore whether the variables vary across conditions approximately as we would expect degrees of anxiety, stress or happiness to do. Nunnally and Bernstein also propose that construct validity obtains if "the supposed measure(s) of the construct *behave as expected* (Nunnally and Bernstein 1994, 90, italics in original).

2.3 Support for the proposition

In this section I will defend the proposition that psychologists with an interest in subjective measures of wellbeing operate within the psychometric approach. The proposition is *prima facie* plausible in light of the fact that the science of happiness is best seen as a branch of personality psychology, which has always been intimately linked with psychometrics. The proposition is further supported by its explanatory power. In this section, I will argue that it accounts for a number of facts regarding the psychologists and their attitudes.

First of all, the proposition that psychologists are interested in subjective measures of wellbeing helps explain the manner in which the psychologists justify their measures. In order to explore their mode of justification, I rely on the 1999 review article by Diener, Eunkook M. Suh, Richard E. Lucas, and Heidi L. Smith (1999). In response to various methodological concerns, the authors respond:

> These measures do possess adequate psychometric properties, exhibiting good internal consistency, moderate stability, and appropriate sensitivity to changing life circumstances. Furthermore, global reports show a moderate level of convergence with daily mood reports, informant reports, spouse reports, and recall for positive versus

negative life events. People who score high on global life satisfaction are less likely to attempt suicide and to become depressed in the future (Diener et al. 1999, 277-278).[4]

In the opinion of these authors, it appears that there are several lines of argument that converge on the conclusion that subjective measures of wellbeing are adequate. A particularly important reason is that subjective measures are positively correlated with phenomena that we would a priori expect to be positively correlated with happiness (such as spouses' assessments of happiness), and negatively correlated with phenomena that we would a priori expect to be negatively correlated with happiness (such as suicide rates).

The reasoning exhibited by Diener et al. (1999) fits the schematic picture painted by Nunnally and Bernstein (1994) very well. First, Diener et al. have identified a set of "observable" variables – in this case, among other things, spouse reports and the absence of suicides – that they take to reflect the same construct as self-reported happiness. Second, Diener et al. have explored the degree to which these variables intercorrelate (positively and negatively) with the self-reports. Third, because all these variables (supposedly) correlate positively, Diener et al. conclude that they all are measures of the same construct. Moreover, Diener et al. do check whether the construct "behave as expected." When they argue that self-reports exhibit "appropriate sensitivity to changing life circumstances" what they mean is that the measure varies across conditions more or less in the expected manner.

The fact that the schematic picture outlined in Kline (1998) and Nunnally and Bernstein (1994) coheres well with the argument presented by Diener et al. (1999) – along with the reference to "adequate psychometric properties" – strongly supports the contention that psychologists defending subjective measures operate within the psychometric approach to measurement. By contrast, it is clear that the methods used by Diener et al. (1999) do not cohere with the

[4] References have been omitted.

measurement-theoretic approach, with its emphasis on observable orderings and representation theorems (see section 3.2 below).

The contention that psychologists operate within the psychometric approach to measurement also explains why it is natural for psychologists to answer questions about wellbeing using questionnaires. Questionnaires have been accepted as a tool for mental testing since the very beginning of personality psychology. Moreover, the historical background explains why psychologists think it is justified to address issues of wellbeing by using self-reports. It is not that they unquestioningly believe that people are capable of accurately revealing the degree of happiness they enjoy. The use of self-reports is supported by a great deal of previous research, which is interpreted as showing that self-report questionnaires are as valid as any other tool.[5]

This history also serves to explain why the economists' approach to welfare measurement tends to strike psychologists as inadequate. First, for somebody trained in the tradition of personality psychology, the use of people's income as a proxy for their welfare is likely to appear as hopelessly indirect. This is so whether welfare is taken to mean degree of desire-fulfillment, happiness or satisfaction. Moreover, so long as economists fail to develop arguments along the lines discussed in the previous section – involving questions of validity and reliability – in favor of their measures, psychologists are likely to consider those measures – whether GNP, consumer surplus, or equivalent variation – unproven.

Furthermore, the contention that psychologists operate within the psychometric approach helps us identify some assumptions that psychologists do not make (or at least, do not need to make). For example, proponents of subjective measures need not assume that individuals in general behave so as to maximize their happiness. The psychometric approach makes no assumptions at all about motivation, so the claim that people in general fail to maximize happiness does not constitute

[5] See Angner (manuscript) for an extended discussion of psychologists' efforts to establish the validity of their measures.

an argument against subjective measures of wellbeing. Moreover, the psychologists do not assume that individuals always know how happy they are, or that they always truthfully reveal their perceived happiness. Even if the psychologists want to rely on self-reports, everything they need to assume is that such reports are sufficiently correlated with other measures of the construct. Thus, the claim that people do not always know whether they are happy, or the claim that they do not always speak the truth, does not in itself constitute an argument against the psychologists' project. Finally, the psychologists need not assume that introspection is a reliable method to assess one's degree of happiness. As far as the psychologists' are concerned, individuals' self-reports may well be based on careful observation of their own behavior. Thus, a general skepticism about the reliability of introspection does not constitute an argument against subjective measures.

3. The economists' approach to measurement

In this section, I explore what makes economists think that they can measure wellbeing or welfare. The main proposition defended in this section is the claim that economists who attempt to measure welfare operate within the measurement-theoretic approach. This proposition, I claim, is supported by the fact that contemporary efforts to measure wellbeing take place within the framework of ordinalism – the view that utility as a psychological entity can be dispensed with, and that the concept of preference can be used as a primitive instead (Mandler 1999, 5) – in combination with the fact that the measurement-theoretic approach is strongly linked to the ordinalist tradition. The proposition also has a great deal of explanatory power. In particular, it helps explain why economists tend to be so critical of psychologists' efforts to measure happiness.

3.1 Ordinalism

Efforts by economists to measure welfare or wellbeing also go back at least until the early 20th century. It is convenient to take as our starting point the work of A. C. Pigou, who is often described as the father of welfare economics. In *The Economics of Welfare* ([1920] 1952), Pigou was explicit about his desire to develop a useful measure of welfare. In his own words: "The goal sought is to make more easy practical measures to promote welfare – practical measures which statesmen may build upon the work of the economist" (Pigou 1952, 10). For Pigou, by the way, questions of welfare were not incidental to the project of economics. Indeed, he explicitly identifies economic welfare with the subject matter of economics (Pigou 1952, 11).

Pigou and other early neoclassical economists explicitly built their economics on the foundation of hedonic psychology, that is, an account of individual behavior according to which individuals seek to maximize utility, where utility was taken to be some mental state like "happiness" or "pleasure." In this they followed Jeremy Bentham, who wrote: "Nature has placed mankind under the governance of two sovereign masters, *pain* and *pleasure*.... They govern us in all we do, in all we say, in all we think" (Bentham [1823] 1996, 11, italics in original). Thus, William Stanley Jevons claimed: "Pleasure and pain are undoubtedly the ultimate objects of the Calculus of Economics. To satisfy our wants to the utmost with the least effort ... in other words, to *maximize pleasure*, is the problem of Economics" (Jevons [1871] 1965, 37). The focus on conscious experience was, of course, shared by welfare economists. As A. C. Pigou put it, "the elements of welfare are states of consciousness and, perhaps, their relations" (Pigou 1952, 10). Because utility was taken to be a matter of conscious experience, the foundations of economics – including the principles of hedonic psychology – were often defended on the basis of their introspective self-evidence (see e.g. Cairnes 1888, 88-90).

However, a major transition took place between the 1930's and 1950's (Mandler 1999, 8). Postwar neoclassical economists objected to the notion that economics should make reference to conscious states, and simultaneously rejected the idea that introspection was a scientifically acceptable means to explore such states. Many of these authors were quite clearly inspired by the methodological strictures of logical positivism in philosophy, behaviorism in psychology, and operationalism in physics. Moreover, at least some economists appear to have grown disappointed with the meager results of accepted economic practice in terms of theories with predictive power (see T. W. Hutchison 1938, 132). Thus, postwar neoclassical economists set out to put their discipline on firmer methodological ground, and at the same time to improve the predictive power of their theories. They did so by rejecting hedonic psychology and instead emphasizing that which can be "objectively observed."

The same postwar economists also believed that references to hedonic states were unnecessary (Mandler 1999, 67). As Lionel Robbins, author of the spectacularly influential *An Essay on the Nature and Significance of Economic Science* ([1932] 1984) wrote, neoclassical economic theory has no "essential connection with psychological hedonism, or for that matter with any other brand of *Fach-Psychologie*" and "is capable of being set out and defended in absolutely non-hedonistic terms" (Robbins 1984, 85). In Robbins' view, economics is based instead on the notion of *relative valuation* or *preference*. As he writes: "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" (Robbins 1984, 85-86).

According to the consensus that emerged after WWII, therefore, the hedonic foundations of early neoclassical economics can be dispensed with, and instead the concept of preference can be used as a primitive. According to some accounts – like the Revealed Preference Theory of Paul Samuelson (1948) – preferences are identified with choices. According to others – like that of Robbins – preferences are not identified with choices (see Robbins 1984, 87-88), but are nevertheless closely linked. In either case it is assumed that preferences can be directly inferred from choices. Both assume that choices satisfy certain consistency conditions (e.g. Robbins 1984, 92). In the postwar view, it remains legitimate to talk about utility, but only as an index of preference satisfaction. Thus, "a utility function carries no psychological meaning or interpretation and serves merely as a mathematical convenience" (Mandler 1999, 5). In welfare economics, the utilitarian welfare criterion, which requires the aggregation of utilities, gave way to the Pareto criterion, which does not (Mandler 1999, 6).

The establishment of ordinalism, as Mandler refers to the postwar consensus, coincided with a shift from mental state accounts of wellbeing (according to which wellbeing is a matter of happiness, satisfaction, or the like) to preference satisfaction accounts (according to which wellbeing is a matter of preference satisfaction). As the historical background suggests, this shift was driven primarily by concerns about what could legitimately be the object of scientific study. An uncharitable reading of this episode would suggest that the economists of the time were like the drunken man who looked for his keys where the light was, rather than where he had lost them. A more charitable interpretation, however, would suggest that the economists accepted some view like that of Griffin and Korsgaard (see section 1), according to which accounts of wellbeing can be judged in part on the grounds of whether they permit the development of adequate measures.

Ordinalism has had a profound impact on the development of modern economics. Since the publication of *The Nature and Significance of Economic Science*, a number of economists have aspired to provide methodological underpinnings for talk about preferences and for the measurement of utility. No doubt, typically the goal was both to render economic method more in line with contemporary methodological strictures, and to improve the predictive accuracy of the theory. Their aspirations assume, like Robbins, that preferences can be unambiguously identified by studying agents' choices. In the next section, I discuss one particularly influential attempt to provide foundations for preference and utility measurement.

3.2 Measurement theory as the foundation for preference and utility

The most sophisticated attempt to provide solid methodological foundations for talk about preference and utility measurement is based on the theory of measurement. Conceptually, of course, in this view the measurement of utility is a specific instance of measurement *tout court*. Historically, however, the two emerged in conjunction. Indeed, the development of the theory of measurement was to a significant degree influenced by problems associated with the measurement of utility (see Krantz et al. 1971, 9). The theory behind the measurement of utility was developed by John von Neumann and Oskar Morgenstern (1944), Leonard J. Savage ([1954] 1972) and others,[6] drawing on Frank P. Ramsey ([1931] 1988). The theory of measurement, in its modern shape, was first articulated by Dana Scott and Patrick Suppes (1958) but received what may be its canonical statement in the three-volume *Foundations of Measurement,* the first volume of which appeared as David H. Krantz, R. Duncan Luce, Patrick Suppes, and Amos Tversky (1971). Here, I also rely on the retrospective article "From Indices to Mappings: The representational approach to measurement" by Krantz (1991).

In their introductory chapter, Krantz et al. (1971) offer a concise statement of the purpose and nature of measurement as they see it. They write:

> When measuring some attribute of a class of objects or events, we associate numbers (or other familiar mathematical entities, such as vectors) with the objects in such a way that the properties of the attribute are faithfully represented as numerical properties (Krantz et al. 1971, 1).

Thus, at one level of abstraction, measurement is the process of attributing numbers to object so as to represent some relevant property of the objects at hand. More specifically, and in more technical language:

---

[6] Cf. Krantz et al. (1971, 9)

> From this standpoint, measurement may be regarded as the construction of homomorphisms (scales) from empirical relational structures of interest into numerical relational structures that are useful. Foundational analysis consists, in part, of clarifying (in the sense of axiomatizing) assumptions of such constructions (Krantz et al. 1971, 9).

In what follows, I will try to explain what this means, and how it applies to the measurement of utility.

It helps to think of measurement in the context of an actual example, so let us follow Krantz et al. (1971) and consider the case of length measurement. In their words:

> Suppose that we have a set of straight, rigid rods whose lengths are to be measured. If we place the rods *a* and *b* side by side and adjust them so that one is entirely beside the other and they coincide at one end, then either *a* extends beyond *b* at the other end, or *b* beyond *a*, or they appear to coincide at that end also. We say, respectively, that *a* is longer than *b*, *b* is longer than *a*, or that *a* and *b* are equivalent in length. For brevity, we write, respectively, $a \succ b$, or $b \succ a$, or $a \sim b$. Two or more rods can be *concatenated* by laying them end to end in a straight line, and so we can compare the qualitative length of one set of concatenated rods with that of another by placing them side by side, just as with single rods. The concatenation of *a* and *b* is denoted $a \circ b$ and the observation that *c* is longer than $a \circ b$ is denoted $c \succ a \circ b$, etc. Many empirical properties of length comparison and of concatenation of rods can be formulated and listed, e.g., $\succ$ is transitive; $\circ$ is associative; if $a \succ b$, then $a \circ c \succ b$; etc. (Krantz et al. 1971, 2).

The basic idea is the following. A set A of objects, in this case a set of rods, can be ordered e.g. with respect to length. We can figure out how various rods are related to each other with respect to length by applying the simple operation described above. This ordering of rods will as a matter of fact satisfy a number of conditions, like transitivity. These conditions can be identified by empirical study, and, incidentally, are most conveniently expressed as a set of axioms (cf. Krantz et al. 1971, 6). The axioms can be seen as a set of empirical laws (Krantz et al. 1971, 13). Thus, “fundamental measurements are based on certain qualitative physical laws” (Krantz 1972, 1428).

The example illustrates what Krantz et al. mean by an *empirical relational structure*.[7] A *relational structure* is a set of objects along with relations on that set (cf. Krantz et al. 1971, 8). In this case, we have a set (which we will call A) of rods. Moreover, we have two relations of interest: $\succ$, which is a binary relation, and $\circ$, which is ternary, holding between a, b, and $c = a \circ b$ (Krantz et al. 1971, 8). Such an empirical relational structure is referred to as $\langle \mathrm{A}, \succ, \circ \rangle$. An empirical relational structure contrasts with a *numerical relational structure*, which is a set of mathematical objects like numbers along with relations on that set.

Given an empirical relational structure, we want to assign numbers $\phi(a)$, $\phi(b)$, etc. to rods a, b, etc. in such a way that the following two conditions are satisfied (see Krantz et al. 1971, 5). First, we require that the number assigned to a be greater than the number assigned to b just in case a in fact is longer than b. That is, $\phi(a) > \phi(b)$ if and only if $a \succ b$. Second, we require that the numbers assigned be additive with respect to concatenation. That is, $\phi(b \circ c) = \phi(b) + \phi(c)$.

If we succeed in assigning numbers such that these two conditions are satisfied, the function $\phi(\cdot)$ is a *homomorphism* from the empirical relational structure into a numerical relational structure $\langle \Re, >, + \rangle$. This means that $\phi(\cdot)$ takes elements of A into the set $\Re$ of real numbers in such a way that the corresponding relationships are preserved. That is, $a \succ b$ if and only if $\phi(a) > \phi(b)$, and $c = a \circ b$ if and only if $\phi(c) = \phi(a) + \phi(b)$. If a homomorphism $\phi(\cdot)$ is one-one, we say that it is an *isomorphism*. This will occur if no two rods are equal with respect to length, so that each rod gets assigned a unique number. Especially in the empirical literature, homomorphisms are often referred to as *scales*. This explains the view of Krantz et al. (1971, 9) that the process of measurement can be

[7] Krantz (1972) refers to them as *measurement structures*, and Krantz (1991) as *qualitative structures*.

seen as the process of constructing homomorphisms, or scales, from empirical to numerical relational structures.

Not every empirical relational structure will allow the construction of a homomorphism, however. As a result, we will want to know what conditions must be satisfied by the empirical relations $\succ$ and $\circ$ on A such that a function $\phi(\cdot)$ with desired properties can be constructed (cf. Krantz et al. 1971, 6). As Krantz et al. write: "A measurement procedure certainly is not adequately understood if it depends on properties that are not explicitly recognized" (Krantz et al. 1971, 6). At this point, it is possible to explore formally what axioms are necessary and sufficient for it to be possible to construct a function $\phi(\cdot)$ that satisfies the properties identified above (Krantz 1971, 8). The answer to a question such as this one is given by a representation theorem, which "asserts that if a given relational structure satisfies certain axioms, then a homomorphism into a certain numerical relational structure can be constructed" (Krantz et al. 1971, 9).

The measurement-theoretic point of view aspires to be – in a certain sense – ontologically non-committal. When measurement theorists talk about length, hunger, frustration, risk aversion, and so on, Krantz notes, it may seem as if they "introduce ontological presuppositions" and presuppose the existence of such things (Krantz 1991, 3). However, he argues, in actual practice measurement theorists do not need to take a position on the issue of the ontological status of such entities, since everything they need to assume is the existence of an empirical ordering that satisfies certain conditions. As Krantz puts it, "in most cases it seems that one expects at least a useful sort of ordering of objects or situations or organisms or social entities" (Krantz 1991, 3). In order to justify talk about length, on this view, everything we need to assume is that there exists an empirical structure of a set of rods and relations that satisfy certain conditions. It is important to notice, by the way, that measurement theory is not an attempt to provide operational definitions of theoretical terms. Krantz et al. write that to treat "indirect measures" as "objective definitions of unanalyzed concepts" is a temptation that has to be resisted (Krantz et al. 1971, 32).

The theory of measurement may seem too abstract to shed any light on real life examples such as the measurement of utility (cf. Krantz et al. 1971, 9). However, as we will see, according to measurement theorists the measurement of utility is a straightforward application of the account described above. In brief, the assumption underlying the measurement of utility is that a choice structure – a set X of options and a choice relation R on X – is an empirical relational structure satisfying certain axioms. Though there are different ways to approach the topic, typically X is the set of all possible acts or bundles, and R is a binary relation such that aRb means that a is chosen over b in a pair-wise choice.

Krantz (1991, section 3) discusses the example of utility quite explicitly. He writes:

> Since 1960, there seems to have been general agreement concerning two main points about the measurement of utility. First, the empirical ordering underlying utility is determined by actual choices; that is, the choice of one act over others is represented by a utility assigned to the chosen act that is higher than the utilities assigned to the other acts (Krantz 1991, 28).

Krantz goes on: "The first of these points reflects the view that it is actual choices that are the most trustworthy and most important data of a behavioral science" (Krantz 1991, 28). Typically, of course, there is an assumption that choices reflect preferences over the various options. At any rate, as Krantz notes, "most utility theories cling to the idea that the ordering is based on observation of choice behavior" (Krantz 1991, 28-29).

In order for the representation theorem to work, the measurement theorist assumes that the choice structure $\langle X, R \rangle$ satisfies some set of conditions. Either one of several different sets of axioms will do the trick. However, there are some conditions that are shared by all axiomatizations of utility. As Krantz et al. (1971, 21-22) point out, transitivity is a *necessary* condition, in the sense that it is mathematically necessary for the representation theorem to work. This claim is proved by the following simple argument. If aRb and bRc, the fact that $\phi(\cdot)$ is a homomorphism implies that $\phi(a) > \phi(b)$ and $\phi(b) > \phi(c)$. Thus, $\phi(a) > \phi(c)$, and since $\phi(\cdot)$ is a homomorphism, aRc. QED.

Just like in the case of the measurement of length, the axioms are seen as empirical (descriptive) laws, in this case, laws of choice. Thus, for instance, the axioms articulated by von Neumann and Morgenstern (1944) "constitute a set of qualitative laws for 'rational' decisions among risky options" (Krantz 1972, 1428). Very often, of course, these axioms are treated as normative laws of rational choice. In the present context, however, the normative status of the axioms is irrelevant. The point here is that the representation theorem requires that they be true descriptive laws. That is, for the theorem to apply the choice structure must *in fact* satisfy the axioms, viz. that the axioms must be *true* of the empirical relational choice structure.

Although the theory of measurement is a complicated affair, and though the presentation in this subsection is brief, we can already see how the measurement of utility is supposed to relate e.g. to the measurement of length. Instead of a set of rods we have a set of bundles or acts. Instead of an ordering determined by comparisons of rods placed side-by-side, we have an ordering determined by the choices of some agent. In order to allow the construction of a representation theorem, we need to identify a set of axioms, which can be seen as empirical, descriptive laws of choice. One of these laws is transitivity, that is, if *a* is longer than *b* and *b* is longer than *c*, then *a* is longer than *c*. Just like in the case of length, measurement theory is supposed to allow us to remain agnostic about the existence of utilities (and preferences, presumably). Thus, the measurement theorist claims to provide methodological foundations for talk about utility and preference without "pretending to look into the head of the agent." All the measurement theorist assumes is that choices determine an empirical relational structure with the appropriate properties, that is, that actual choices satisfy the appropriate axioms. Note the importance of the consistency (transitivity) condition. If this condition is not in fact satisfied, measurement theory gives us no grounds whatsoever for constructing a utility function on the set of available options.

3.3 Support for the proposition

In this section I will defend the proposition that economists interested in welfare measurement operate within the measurement-theoretic approach. This proposition, I claim, is supported by the fact that contemporary economists remain strongly ordinalist, in combination with the fact that the measurement-theoretic approach is strongly linked to the ordinalist tradition. The link is suggested, among other things, by the fact that the development of axiomatic utility theory – which remains absolutely central to modern economics – was inspired in part by problems of measurement (Krantz et al. 1971, 9). The proposition is further supported by its explanatory power. In this section, I will argue that it sheds light on economists' efforts to measure welfare.

First of all, the proposition helps explain the type of data that economists use in order to construct their measures of welfare. Thus, it explains why economists favor measures that take data about economic transactions – like market choices – as their starting point. The central idea behind most attempts by economists to measure wellbeing is neatly expressed by Richard E. Just, Darrell L. Hueth, and Andrew Schmitz (1982). They begin by pointing out that utility is not observable, and add:

> In most practical situations the applied welfare economist can, at best, observe income and consumption decisions at various prices and then, on the basis of these economic transactions, try to compute some money-based measure of welfare effects (Just et al. 1982, 69).

Though the economic approaches to the measurement of wellbeing differ in many ways, almost all of them rely on observable transactions in order to represent utility. In particular, the proposition helps account for economists' skepticism regarding questionnaire studies.

This skepticism regarding data about anything other than market-based choice is well documented. In a discussion of dominant attitudes among economists, Amartya Sen comments:

> Choice is seen as solid information, whereas introspection is not open to observation.... Much of economic theory seems to be concerned with strong, silent men who never speak! One has to sneak in behind them to see what they are doing in the market, etc., and deduce from it what they prefer, what makes them better off, what they think is right, and so on (Sen 1982, 9).

Similarly: "Much of the empirical work on preference patterns [and therefore welfare] seems to be based on the conviction that [non-verbal] behaviour is the only source of information on a person's preferences" (Sen 1982, 71). These attitudes can, at least to some extent, be accounted for by pointing out that economists work within the measurement-theoretic approach to measurement, which puts a very strong emphasis on observable orderings like that imposed on a set of alternatives by an agent's choices.

Second, the proposition sheds some light on the manner in which economists go about defending their measures. As we saw in section 2.3, psychologists are sometimes curious about the fact that economists do not try to establish the validity and reliability of their measures in the manner of the psychometric approach. In the spirit of measurement theory, however, what you need to do to show that a given measure represents utility (or welfare, which comes down to the same thing) is to show that it is a homomorphism, which is a rather different task. Economists tend to assume that the use of a given measure has been justified when it has been shown that it is based on market choices assumed to satisfy the relevant axioms, in conjunction with a formal proof that shows that the measure is an index of preference (i.e. a utility function). Because this procedure, if successful, in fact establishes that the measure is a homomorphism, the measurement-theoretic approach helps explain why the procedure is followed.

Third, the proposition suggested here helps explain economists' strict adherence to preference satisfaction accounts of wellbeing. The adoption of a preference satisfaction account of wellbeing was part of the project of showing that references to mental states were unnecessary, and that the concept of preference would suffice to provide foundations for economic theory. Economists like Robbins certainly considered the claim that preference satisfaction can be measured a major consideration in favor of preference satisfaction accounts of wellbeing. It is plausible to assume that this sentiment remains accurate today, though it is less explicit than in Griffin and Korsgaard.

Finally, the proposition goes a long way toward accounting for economists' skepticism toward subjective measures. The belief that there is no observable ordering that can be used as a basis for the measurement of e.g. happiness accounts for the belief that it cannot be measured. I imagine these attitudes go a long way toward accounting for Beckerman's criticism. He and many others, following the measurement-theoretic approach, believe that a proper measure is based on a behavioral, observational ordering, and is skeptical of any measure that has not been shown to be a homomorphism. In the next section I will argue that the proposition can be used to spell out what I have called the measurement-theoretic argument against subjective measures of wellbeing.

4. The measurement-theoretic argument

Assuming that contemporary economists operate within the measurement-theoretic approach, it is a straightforward matter to identify the central premisses of the argument against subjective measures. The background reviewed above suggests an argument along the following lines. First:

(1) Measurement requires the existence of an observable ordering.

As we saw above, this is a central assumption of the theory of measurement. Indeed, if the relation of interest fails to satisfy e.g. transitivity, then the theory gives us no grounds for proceeding with assigning numbers.[8] Second:

(2) The (observable) choices of economic agents constitute such an ordering.

This assumption, in some form or other, was endorsed both by Robbins, Samuelson, and the measurement theorists. In the language of the latter, assumption (2) is equivalent to the following:

(2') The (observable) choices of economic agents satisfy the axioms of rational choice theory.

Thus, assumption (2') says that agents' choices in fact satisfy the "laws" of rational choice. Moreover, an assumption that is shared by all the authors discussed in the previous section is the following:

(3) The (observable) choices of economic agents reflect their preferences, in the sense that *A* is chosen over *B* just in case *A* is preferred over *B*.

This assumption, though implicit, is clearly present both in Robbins, Samuelson, and Krantz et al. Its epistemological status differs, of course. Samuelson takes the assumption to be a definitional truth rather than an empirical one. By contrast:

(4) There is no corresponding ordering in the case of the measurement of happiness, satisfaction, and so on.

Thus, (4) says that there is no observable ordering on which measurement can be based. This claim appears to be completely unargued, but it is required for the argument to be at all effective.

Together, these claims help us understand what I call the measurement-theoretic argument, and in particular the claim that happiness cannot be measured whereas preference satisfaction can.

---

[8] The claim that an observable ordering is a necessary condition for measurement is further strengthened by the Krantz passage quoted below (section 5).

Claims (1) and (4) together imply that degrees of happiness cannot be measured. Meanwhile, claims (1) through (3) imply that degrees of preference satisfaction can.

5. Is the argument convincing?

Assuming that I have correctly identified the assumptions that go into the measurement-theoretic argument against subjective measures, I now proceed to examine its plausibility. I will focus the claim that mental states like happiness do not permit the development of adequate measures whereas preference satisfaction does. As we saw above, this claim is based on the notion that measurement requires the existence of an observable ordering.

First, note that many economists argue that wellbeing consists in the satisfaction of those preferences that the agent would have under some specified counterfactual conditions (see e.g. Mongin and d'Aspremont 1998, 397). Yet, there is some serious tension – not to say contradiction – between adopting the measurement-theoretic approach to measurement, and adopting a preference-satisfaction account of wellbeing according to which what matters are counterfactual preferences. While such accounts may be more plausible as accounts of wellbeing, the preference ordering that you would have under some counterfactual conditions – or, the choices that you would make under those conditions – are unobservable by design. Thus, if wellbeing is understood to be a matter of the satisfaction of ideal preferences, there is no observable ordering that could serve as the basis for measurement.

So let us instead consider the preferences that you actually have. Do these preferences imply the existence of an observable ordering? In fact, Krantz (1991) discusses this issue.[9] Krantz is impressed by empirical results that seem to suggest that people systematically violate the axioms of rational decision-making. Under the heading "The Myth of Utility" (Krantz 1991, 28), Krantz cites a

[9] A similar conclusion has been drawn by Daniel Kahneman, Ilana Ritov, and David Schkade (1999).

series of empirical results, and notes: "Choice does indeed depend on the method of testing ... and depends especially on how options are framed" (Krantz 1991, 32). What Krantz calls the "myth of preference" is, he says, "linked to the behavioral assumption that 'preferences' are 'revealed' by choices or 'elicited' by presentation of suitable options and to the mathematics of maximization" (Krantz. 1991, 35). In Krantz's view, since agents' choices fail to conform to assumptions of rational choices theory, these agents do not have preferences; hence the use of the term "myth." This view, if correct, would pose a serious obstacle to the notion that welfare is a matter of the satisfaction of one's preferences.

From our vantage point, if anything, the case for the truth of the axioms of rational choice theory is even weaker than it was when Krantz wrote his retrospective in 1991. Many different researchers claim to have found evidence to the effect that people's choices, to a very significant extent, reflect incidental aspects of the decision situation rather than a stable, consistent preference ordering. As Matthew Rabin (2002) puts it: "A lot of decisions are so sensitive to the framing or context of the choice set that it is difficult to associate these decisions as coming from framing- or context-free preferences on those choice sets" (Rabin 2002, 662). Similarly, Tversky writes that "if different methods of elicitation give rise to different choices," then "it is difficult to defend the proposition that a person has a well-defined preference order (or equivalently a utility function)" (Tversky 1996, 189).

So long as we insist on using the measurement-theoretic approach to measurement, therefore, it is hard to maintain that degrees of preference satisfaction can be measured. If wellbeing is a matter of ideal preferences, it is difficult to argue that there is an observable ordering that can serve as a basis for measurement. If wellbeing is a matter of actual preference satisfaction, it is equally difficult (but for other reasons) to argue that there is an observable ordering that can serve as a basis for measurement. Either way, then, the central claim underlying the economic argument against subjective measures appears to be false. If subjective measures fail for the reason identified by Beckerman (1974) and others, economic measures fail too, and for much the same reason.

6. Discussion

In this section, I summarize the conclusions from earlier sections, explore some implications, and consider some potential counterarguments. In brief, I have argued that psychologists with an interest in the measurement of wellbeing operate within the psychometric approach, whereas economists work within the measurement-theoretic one. This fact, I have maintained, sheds a great deal of light on psychologists' and economists' efforts to measure wellbeing. In particular, I have argued, the contention that economists operate within the measurement-theoretic approach helps us to spell out what I have called the measurement-theoretic argument. The analysis of this argument reveals that it relies critically on an empirical premiss that is increasingly hard to defend, no matter whether we take wellbeing to be a matter of actual or counterfactual preferences. Thus, the psychological measures cannot be dismissed quite as easily as some critics have envisioned.

It does not follow, however, that degrees of happiness are easier to measure than degrees of preference satisfaction. In my view, the emphasis on measurement-theoretic considerations (and the absence of observable orderings) has served to obscure the real problems that afflict subjective measures, and which are quite real. The first concern is that the entire psychometric approach is misguided. If so, it should be possible to develop a general argument against the psychometric approach across the board, whether for the purposes of measuring intelligence, optimism/pessimism, or happiness. This is perhaps unlikely in light of the fact that – as suggested by Johnson (2001, 11316) – the psychometric method is a fairly straightforward application of the hypothetico-deductive method of hypothesis testing (cf. Earman and Salmon 1992, 44 *ff*).

The second concern, which is more serious, is that the psychometric approach is acceptable in principle, but that the happiness and satisfaction constructs, and their measures, have not been adequately validated. If so, it should be possible to articulate an argument to the effect that the correlations identified by the psychologists are not strong enough to warrant the their inferences (although other constructs like intelligence may in fact have been adequately validated). Even the psychologists themselves are aware of problems of this general kind. As Diener et al. (1999) write: "SWB values may change depending on the type of scales used, the order of items, the time-frame of the questions, current mood at the time of measurement, and other situational factors" (Diener et al. 1999, 278).[10] In the light of phenomena like these, Norbert Schwarz and Fritz Strack (1999) write: "Reports of subjective well-being (SWB) do not reflect a stable inner state of well-being. Rather, they are judgments that individuals form on the spot, based on information that is chronically or temporarily accessible at that point in time, resulting in pronounced context effects" (Schwarz and Strack 1999, 61).

The third concern, and the most serious, is that psychologists rely on the wrong concept of wellbeing. While accounts according to which wellbeing is a matter of happiness (or the like) have been popular in the past, they are now considered so implausible that there is not a single living philosopher (to the best of my knowledge) who is willing to defend it. As L. W. Sumner (1996) puts it:

[10] References have been omitted.

> Time and philosophical fashion have not been kind to hedonism. Although hedonistic theories of various sorts flourished for three centuries or so in the congenial empiricist habitat, they have all but disappeared from the scene. Do they now merit even passing attention, for other than nostalgic reasons? (Sumner 1996, 83).[11]

I am not suggesting, of course, that philosophical issues be settled by a simple vote. Nevertheless, the fact that no living philosopher appears willing to defend a particular position must be considered unusual; no matter how implausible a point of view, there are typically philosophers willing to accept it. If wellbeing has little to do with happiness, then a measure of happiness (no matter how valid) need not reflect wellbeing in any interesting way.

Whether or not these three concerns are real problems for subjective measures, they certainly deserve further attention. Yet, as should be clear, the measurement-theoretic argument has no bearing on these issue. Indeed, by focusing so strongly on measurement-theoretical considerations, and the absence of observable orderings, economists have obscured the real problems, which have nothing to do with the existence of such orderings.

Meanwhile, the discussion appears to paint a bleak picture for the future of economic measures: if Krantz is correct, utility is a myth, and talk about preferences has no foundation. However, it seems that Krantz's conclusion is too strong. It is possible to maintain a belief in the existence of preferences even in the face of the empirical evidence discussed above. Consistency of choice is not a necessary condition for having preferences. It is quite possible to maintain that people do have preferences, but that preference is related to choice in some less reliable way. This view would have a number of important implications. First, the exact relationship between choices and preferences would be an empirical question, which would need to be settled by scientific

---

[11] While Sumner himself intends to revive a version of hedonism, he introduces the additional requirement that the happiness be *authentic* (Sumner 1996, 172), which makes it very different from what contemporary psychologists have in mind.

research. Second, although choices would still serve as evidence of preference, it would be illegitimate to infer statements about preference s directly from statements about choices.

For the defenders of economic measures of wellbeing, this strikes me as a reasonable attitude in the face of empirical results like those discussed above. It can still be claimed that choices can be used to develop a rough idea of the level of preference satisfaction of a given individual, and that the amount of resources available to an agent can serve as an indicator of her wellbeing. Surely, if preferences exist at all, choices can be used as *some* evidence of preferences. The use of choice data as evidence for claims about preference does not, of course, preclude other kinds of evidence. For a long time, Amartya Sen has urged economists to allow e.g. verbal reports. He writes: "That behaviour is a major source of information on preference can hardly be doubted, but the belief that it is the only basis of surmising about people's preferences seems extremely limiting" (Sen 1982, 71). He notes that there are problems associated e.g. with "the questionnaire method," but adds that there are problems with using choice behavior as well (Sen 1982, 71).

This move would, however, have some interesting implications for the measurement of welfare. If economists reject the notion that measures of welfare can be based on an observable, behavioral ordering, the standard arguments for their validity have to be replaced by other arguments. These arguments cannot, obviously, be based on the measurement-theoretic approach to measurement. It is difficult to guess what such arguments would look like, or how successful they would be. However, a fair guess is that they would follow the tracks laid down by the psychometric approach. After all, in its terms preferences would be a latent variable, or a "construct," and may be measured according to the methods psychologists have developed for this task.

By way of conclusion, I wish to return to the historical themes addressed in the early part of the paper. The historical background reviewed above shows that psychologists and economists started out during the early part of the 20th century with very similar ideas about what constituted wellbeing or welfare (in spite of the fact that they proposed different methods to measure it). Between the 1930's and 1950's, however, economists gradually moved away from their roots in hedonic psychology and shed all references to pleasure, happiness, and the like. While psychologists never abandoned their view that wellbeing is a matter of mental states (see Angner manuscript), economists instead adopted the view that it is a matter of preference satisfaction. In light of this background, it is interesting to ask what, exactly, explains the divergent paths.

One possible suggestion is that the economists, but not the psychologists, shunned pleasure because they were influenced by positivism, behaviorism, and operationalism. This suggestion does quite not ring true, however, since e.g. behaviorism was enormously influential in psychology as well for most of the 20th century. And at any rate, it would have been possible for psychologists to continue talking about pleasure and happiness, even if those terms were defined in terms of behavior patterns alone, like the behaviorists would suggest.

Another suggestion, and a more plausible one, is that the divergence was a function in part of the applications of the research. The economists' research project was driven in large part by a cluster of theoretical problems, such as how to provide foundations for demand analysis (Mandler 1999, 68 *ff.*), which they found could be adequately addressed by ordinalism. Meanwhile, the personality psychologists have remained preoccupied with very practical problems. According to Winter and Barenbaum (1999, 5), "the first objective, self-report 'inventory' purporting to measure a personality characteristic" was developed by a committee set up by the American Psychological Association to predict what soldiers would be vulnerable to "shell shock" as the U.S. entered WWI. To the best of my knowledge, the military, the government, and various corporations have never stopped demanding such tests – in part, no doubt, because they are less susceptible to methodological fashion in philosophy, psychology, and physics – and this may have provided an incentive for psychologists to not shun pleasure like their economist colleagues did.

BIBLIOGRAPHY

Angner, Erik (2005), *Subjective Measures of Well-Being: A philosophical examination*. University of Pittsburgh Doctoral Dissertation.

Angner, Erik (manuscript), "The Evolution of Eupathics: The historical roots of subjective measures of wellbeing."

Bentham, Jeremy ([1823] 1996), *An Introduction to the Principles of Morals and Legislation*. Oxford: Clarendon.

Cairnes, John E. ([1888] 1965), *The character and logical method of political economy,* Second Ed. New York: A. M. Kelley.

Davis, Katharine Bement (1929), *Factors in the Sex Life of Twenty-Two Hundred Women*. New York: Harper & Brothers.

Diener, Ed, Richard E. Lucas, and Shigehiro Oishi (2002), "Subjective Well-Being: The science of happiness and life satisfaction," in C. R. Snyder and Shane J. Lopez (eds.), *Handbook of Positive Psychology*. Oxford: Oxford University Press, pp. 63-73.

Diener, Ed, Eunkook M. Suh, Richard E. Lucas, and Heidi L. Smith (1999), "Subjective Well-Being: Three decades of progress," *Psychological Bulletin* 125: 276-303.

Earman, John and Wesley C. Salmon (1996) "The Confirmation of Scientific Hypotheses," in Merrilee H. Salmon, John Earman, Clark Glymour, James G. Lennox, Peter Machamer, J. E. McGuire, John D. Norton, Wesley C. Salmon, and Kenneth F. Schaffner, *Introduction to the Philosophy of Science*. Englewood Cliffs: Prentice Hall, pp. 42-103.

Griffin, James (1986), *Well-Being: Its meaning, measurement and moral importance*. Oxford: Clarendon.

Gurin, Gerald, Joseph Veroff, and Sheila Feld (1960), *Americans View Their Mental Health: A nationwide interview survey*. New York: Basic Books.

Hamilton, Gilbert V. (1929), *A Research in Marriage*. New York: Albert & Charles Boni.

Hutchison, T. W. (1938), *The Significance and Basic Postulates of Economic Theory*. London: Macmillan.

Jevons, W. Stanley ([1871] 1965) *The Theory of Political Economy*, Fifth Ed. New York: A. M. Kelley.

Johnson, J. A. (2001), "Personality Psychology: Methods," in Smelser and Baltes (2001), Vol. 16, pp. 11313-11317.

Just, Richard E., Darrell L. Hueth, and Andrew Schmitz (1982), *Applied Welfare Economics and Public Policy*. Englewood Cliffs: Prentice-Hall.

Kahneman, Daniel, Ed Diener, and Norbert Schwarz (eds.) (1999), *Well-Being: The foundations of hedonic psychology*. New York: Russell Sage.

Kline, Paul (1998), *The New Psychometrics: Science, psychology and measurement*. London: Routledge.

Korsgaard, Christine M. (1993), "G. A. Cohen: Equality of What? On Welfare, Goods and Capabilities. Amartya Sen: Capability and Well-Being," in Martha Nussbaum and Amartya Sen (eds.), *The Quality of Life*. Oxford: Oxford University Press, pp. 54-61.

Krantz, David H. (1972), "Measurement Structures and Psychological Laws," *Science* 175: 1427-1435.

Krantz, David H. (1991), "From Indices to Mappings: The representational approach to measurement," in Donald R. Brown and J. E. Keith Smith (eds.), *Frontiers of Mathematical Psychology: Essays in honor of Clyde Coombs*. New York: Springer, pp. 1-52.

Krantz, David H., R. Duncan Luce, Patrick Suppes, and Amos Tversky (1971), *Foundations of Measurement,* Vol. I. New York: Academic Press.

Mandler, Michael (1999), *Dilemmas in Economic Theory: Persisting foundational problems of microeconomics*. New York: Oxford University Press.

McAdams, O. P. (2001), "Personality Psychology," in Smelser and Baltes (2001), Vol. 16, pp. 11308-11313.

Mongin, Philippe and Claude d'Aspremont (1998), "Utility Theory and Ethics," in Salvador Barberà, Peter J. Hammond, and Christian Seidl (eds.), *Handbook of Utility Theory*, Vol. I. Dordrecht: Kluwer, pp. 371-481.

von Neumann, John and Oskar Morgenstern (1944), *Theory of Games and Economic Behavior*. Princeton: Princeton University Press.

Nunnally, Jum C. and Ira H. Bernstein ([1967] 1994), *Psychometric Theory*, Third Ed. New York: McGraw-Hill.

Pigou, A. C. ([1920] 1952) *The Economics of Welfare*, Fourth Ed. London: Macmillan.

Ramsey, Frank P. ([1931] 1988) "Truth and Probability," in Peter Gärdenfors and Nils-Eric Sahlin (eds.) *Decision, Probability, and Utility: Selected readings*. Cambridge: Cambridge University Press, 1988, pp. 19-47.

Robbins, Lionel C. ([1932] 1984) *An Essay on the Nature and Significance of Economic Science*, Third Ed. New York: New York University Press.

Samuelson, Paul A. (1948) "Consumption Theory in Terms of Revealed Preference," *Economica,* 15: 243-253.

Savage, Leonard J. ([1954] 1972) *Foundations of Statistics*, Second Revised Ed. New York: Dover.

Scott, Dana and Patrick Suppes (1958), "Foundational Aspects of Theories of Measurement," *The Journal of Symbolic Logic* 23: 113-128.

Sen, Amartya (1982) *Choice, Welfare and Measurement* (Cambridge: Harvard University Press).

Schwarz, Norbert and Fritz Strack (1999), "Reports of Subjective Well-Being: Judgmental processes and their methodological implications," in Kahneman et al. (1999), pp. 61-84.

Smelser, Neil J. and Paul B. Baltes (eds.) (2001), *International Encyclopedia of the Social and Behavioral Sciences.* Amsterdam: Elsevier.

Watson, Goodwin (1930), "Happiness Among Adult Students of Education," *The Journal of Educational Psychology* 21: 79-109.

Winter, David G. and Nicole B. Barenbaum (1999) "History of Modern Personality Psychology Theory and Research," in Lawrence A. Pervin and Oliver P. John (eds.), *Handbook of Personality: Theory and research*, Second Ed. New York: Guilford.