CURRENT TRENDS IN WELFARE MEASUREMENT

Erik Angner, Ph.D.

Date: 20 September 2010

Abstract. The rise of subjective measures of well-being represents at least two important trends in the measurement of welfare or well-being. The first trend, which has already received some attention in the literature, is a shift away from preference-satisfaction accounts of individual well-being and toward mental-state accounts. The second trend, which has gone largely unnoticed, is a shift away from the measurement-theoretic (or representational) approach to measurement and toward the psychometric approach. In this chapter, I will argue that whereas orthodox economic welfare measures are based on the measurement-theoretic approach, subjective measures are based on the psychometric approach. The difference helps explain why subjective measures are based on questionnaire data, while orthodox economic measures are based on observable choices; why proponents of subjective measures validate their measures by establishing construct validity, reliability, and so on, whereas orthodox economists tend to establish that a particular function is a utility function; why orthodox economists' approach to welfare measurement strikes proponents of subjective measures as terribly inadequate, and vice versa; and why subjective measures are based on mental-state accounts, whereas orthodox economic measures are based on preference-satisfaction accounts of well-being. This trend constitutes a radical methodological shift, which is likely to have a significant impact on the shape of welfare economics and on the public policy it informs, and could generate novel and interesting avenues of research.

To appear in John B. Davis and D. Wade Hands (Eds.) *The Elgar Companion to Recent Economic Methodology* (Northhampton: Edward Elgar). Contact information: Erik Angner, Dept. of Philosophy and Dept. of Marketing, Industrial Distribution, and Economics, Humanities Building 414A, University of Alabama at Birmingham, 1530 3rd Avenue South, Birmingham, AL 35294-1260, USA. Email: angner@uab.edu. I am grateful to the editors of this volume for constructive criticism of an earlier draft. Errors remain my own.

1 Introduction

Welfare or *well-being* – I will use the terms interchangeably – is "what we have when our lives are going well for us, when we are living lives that are not necessarily morally good, but good *for us*" (Tiberius, 2006, p. 493). Because the concept of well-being is frequently assumed to capture what is – ultimately and not just instrumentally – good for individuals and for groups, it is also supposed to capture that which we have reason to promote – as an end and not just as a means – in our own lives and in the lives of others (cf. Scanlon, 2000, pp. 108-109). Hence, it is widely assumed that well-being is one (or *the* one) consideration that should serve as an end – and not just a means – for public policy. "A full consideration of taxes, subsidies, transfer programs, health care reform, regulation, environmental policy, the social security system, and educational reform," Daniel T. Slesnick writes, "must ultimately address the question of how these policies affect the well-being of individuals" (Slesnick, 1998, p. 2108).

Given the importance of the concept of welfare, it is unsurprising that the *measurement* of welfare should be a central concern of modern economics – both theoretical and applied – and of public policy analysis (Slesnick, 1998, p. 2108). To develop a working definition of "measurement," I will follow David H. Krantz, R. Duncan Luce, Patrick Suppes, and Amos Tversky:

> 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, Luce, Suppes, & Tversky, 1971, p. 1).

Thus, loosely speaking, *measurement* is the process of assigning numbers or other mathematical entities to objects so as to represent some attribute of those objects; a *measure* is a function from

a set of objects to a set of numbers or other mathematical entities. In the context of welfare measurement, the attribute of interest is welfare; the objects are individuals (in the case of individual welfare measurement) or groups (in the case of social welfare measurement).

Welfare measurement has been central to economics at least since the time of A. C. Pigou, author of *The Economics of Welfare* (1952; 1st ed. 1920). Pigou, who believed that "economic welfare … is the subject-matter of economic science," made it a crucial goal to develop practicable welfare measures (Pigou, 1952, pp. 10-11). Pigou drew a distinction between "total welfare" and "economic welfare," where the latter is "that part of social welfare that can be brought directly and indirectly into relation with the measuring-rod of money" (Pigou, 1952, p. 11). Yet, he proceeded to argue that "qualitative conclusions about the effect of an economic cause upon economic welfare will hold good also of the effect on total welfare" (Pigou, 1952, p. 20), which would entail that for practical purposes the difference between economic and total welfare can be ignored and that measures of economic welfare can be used as proxies for total welfare. These days, orthodox economic welfare measures – including measures of national income, consumer and producer surplus, as well as equivalent and compensating variation (see section 3.1 below) – remain by far the most commonly used measures of welfare.

In recent years, however, so-called *subjective measures of well-being* have increasingly been presented as substitutes for or complements to orthodox economic measures (Campbell, 1976; Diener, 2000, 2006; Diener, Lucas, Schimmack, & Helliwell, 2009; Diener & Seligman, 2004). Subjective measures of well-being – sometimes called *measures of subjective well-being* – are typically based on direct 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, & Feld, 1960, p. 411). Answers to such questions are used to construct

numerical measures of both individual and social well-being (Angner, 2009a, 2009b). Though subjective measures originated in psychology in the 1920's and 30's, increasing numbers of *bona fide* economists are now endorsing their use (Blanchflower & Oswald, 2004; Clark & Oswald, 2002; Frey, 2008; Frey & Stutzer, 2002; Kahneman & Krueger, 2006; Kahneman, Krueger, Schkade, Schwarz, & Stone, 2004a, 2004b; Krueger, 2009; Layard, 2005). Furthermore, policy makers appear to be paying attention. In a recent commencement address, for example, U. S. Federal Reserve Chairman Ben S. Bernanke argued: "Happiness research can be useful for individuals, but it also has implications for policy" (Bernanke, 2010, p. 10).

The rise of subjective measures represents at least two important trends in the measurement of welfare or well-being. The first trend, which has already received some attention in the literature, is a shift away from preference-satisfaction accounts of individual well-being and toward mental-state accounts (Adler & Posner, 2008; Angner, 2008, 2009b, 2010a).[1] Orthodox economic welfare measures are based on preference-satisfaction accounts of well-being, according to which a person is well off to the extent that his or her preferences are satisfied; that is, orthodox economic measures are designed to represent degrees of preference satisfaction. By contrast, subjective measures of well-being are based on mental-state accounts of well-being, according to which a person is well off to the extent that he or she is in some particular subjectively experienced mental state like happiness or satisfaction; that is, subjective measures are designed to represent degrees of happiness, satisfaction, and the like. As one indication of the recent surge in interest in mental states like happiness, consider the fact that the

[1] See Angner (2010b) for more about the conception of individual well-being implicit in literature on subjective measures of well-being. By contrast, subjective measures of well-being are based on the same conception of social well-being as orthodox economic measures (Angner, 2009a).

first edition of the *New Palgrave Dictionary of Economics*, originally published in 1987, does not even have an entry for happiness (Eatwell, Milgate, & Newman, 1998), whereas the most recent (second) edition, published in 2008, dedicates almost six pages to the topic (Durlauf & Blume, 2008).

The shift from preference-satisfaction to mental-state accounts of well-being constitutes a major methodological transition. Since it is possible to be happy and/or satisfied even though one's preferences are not satisfied and *vice versa*, a person who is well off according to the one account is not necessarily well off according to the other. It follows that the individual-welfare criteria proposed in literature on subjective measures of well-being are radically different from those used by orthodox economists, for example, in the two fundamental theorems of welfare economics. Indeed, the simultaneous endorsement of the two individual-welfare criteria would lead to inconsistency. Consider two states of the world x and y such that a person P strictly prefers x to y but is happier in y than in x. Here, the orthodox economic welfare criterion ranks x strictly above y in terms of P's welfare while the welfare criterion advocated in the literature on subjective measures ranks y strictly above x. Hence, the simultaneous endorsement of the two welfare criteria would entail that x is ranked strictly above *and* strictly below y, which is impossible. The fact that subjective and economic measures are based on different conceptions of individual well-being is frequently overlooked, which has generated a great deal of confusion (Angner, 2009b). Because this trend has been discussed elsewhere, however, it will not be my focus here.

The second trend, which has gone largely unnoticed, is a shift away from the measurement-theoretic (or representational) approach to measurement and toward the psychometric approach. In this chapter, I will argue that orthodox economic welfare measures

and subjective measures of well-being are based not just on radically different conceptions of well-being, but on radically different approaches to measurement as well: whereas orthodox economic welfare measures are based on the measurement-theoretic approach, subjective measures are based on the psychometric approach. The difference helps explain why subjective measures are based on questionnaire data, while orthodox economic measures are based on observable choices; why proponents of subjective measures validate their measures by establishing construct validity, reliability, and so on, whereas orthodox economists tend to establish that a particular function is a utility function; why orthodox economists' approach to welfare measurement strikes proponents of subjective measures as terribly inadequate, and *vice versa;* and indeed, why subjective measures are based on mental-state accounts, whereas orthodox economic measures are based on preference-satisfaction accounts of well-being.

The shift from the measurement-theoretic to the psychometric approach to measurement, which I will argue is evident also outside of the literature on subjective measures, constitutes another major methodological transition. Since it is possible to satisfy the strictures imposed by the one approach to measurement without satisfying those imposed by the other, a measure that has been validated in accordance with the one approach has not necessarily been validated in accordance with the other. Indeed, the simultaneous endorsement of the two approaches to measurement would lead to inconsistency. Like in the case of the underlying conceptions of well-being, a failure to notice the difference between the two approaches can generate a great deal of confusion. Although the trend away from the measurement-theoretic and toward the psychometric approach has not yet received much attention, the fact that it constitutes a radical methodological shift means that it is likely to have a significant impact on the shape of welfare

economics and on the public policy that it informs, and could generate thoroughly novel and highly interesting avenues of research.

2 Approaches to measurement

It is widely recognized that there are, broadly speaking, two different approaches to measurement in social and behavioral science (Dawes & T. L. Smith, 1985; John & Benet-Martínez, 2000; Judd & McClelland, 1998; Krantz, 1991).[2] As Krantz (1991) puts it: "One, which may be termed the *psychometric* approach, introduces latent [unobservable] variables to explain behavioral orderings" (Krantz, 1991, p. 2). This approach emphasizes latent constructs, reliability, and construct validity. Krantz continues: "The second ... treats the numerical representation of behavioral orderings axiomatically" (Krantz, 1991, p. 2). This approach emphasizes observable orderings, homomorphisms, and representation theorems. The second approach is sometimes – e.g., in Robyn Dawes and Tom L. Smith (1985, pp. 511-512) – referred to as the *representational* approach. For the reason identified by Dawes and Smith – the fact that all measurement is at bottom about representation – I will call it the *measurement-theoretic* approach. In this section, I will discuss the two approaches in reverse order. We will see that the differences between the two approaches have deep implications for the nature of measurement in social and behavioral science.

2.1 The measurement-theoretic approach

The measurement-theoretic approach was first articulated by Dana Scott and Suppes (1958) but received its canonical statement in the three-volume *Foundations of Measurement*, the first

[2] Denny Borsboom (2005) adds a third approach, *classical test theory*, which I will ignore here.

volume of which appeared in 1971 (Krantz et al., 1971). In order to develop a better idea of the nature of this approach, I will also rely on two articles by Krantz: "Measurement structures and psychological laws" (1972) and "From indices to mappings: The representational approach to measurement" (1991). As we will see, this approach emphasizes observable orderings, homomorphisms, and representation theorems.

Here is how the authors of *Foundations of Measurement* capture the essence of what I have called the measurement-theoretic approach: "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" (Krantz et al., 1971, p. 9). In what follows, I will try to explain what this means. 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, p. 2).

The fundamental idea is that a set A of objects, in this case rods, can be ordered with respect to length by means of an observable operation. That is, it can be determined how various rods are

related to each other with respect to length by applying a series of simple operations like those described above. The ordering of rods will as a matter of fact satisfy a number of conditions, including that of transitivity: if *a* is longer than *b*, and *b* is longer than *c*, it follows that *a* is longer than *c*. These conditions, which can be established by empirical study, can be expressed as a set of axioms (Krantz et al., 1971, p. 6), which in turn can be seen as a set of empirical laws (Krantz et al., 1971, p. 13). Thus, "fundamental measurements are based on certain qualitative physical laws" (Krantz, 1972, p. 1428).

The example illustrates what Krantz et al. mean by a *relational structure:* it is a set of objects along with certain relations on that set (Krantz et al., 1971, p. 8).[3] In this case, we have a set A of rods as well as two relations on that set: $\succ$, which is a binary relation, and $\circ$, which is ternary, holding between a, b, and $c=a\circ b$. The relational structure is referred to as $\langle \mathrm{A}, \succ, \circ \rangle$. The relational structure is *empirical* in the sense that (i) the members of A are observable entities and (ii) the relations between these entities can be established by means of a series of observable operations. 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 can construct a measure of length by assigning numbers $\phi(a)$, $\phi(b)$, etc., to rods a, b, etc., in such a way that two conditions are satisfied (Krantz et al., 1971, p. 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, for all a and b, $\phi(a)>\phi(b)$ if and only if $a\succ b$. Second, we require that the numbers assigned be additive with respect to concatenation; that is, for all b and c, $\phi(b\circ c)=\phi(b)+\phi(c)$. If we succeed in assigning numbers

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

such that these two conditions are satisfied, the function $\phi(\cdot)$ is a *homomorphism* from the empirical relational structure into a numerical relational structure $\langle \mathbb{R}, >, + \rangle$. This means that $\phi(\cdot)$ takes elements of A into the set $\mathbb{R}$ 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 $\phi(c) = \phi(a) + \phi(b)$.[4] Especially in the empirical literature, homomorphisms are often referred to as *scales*. This explains the view of Krantz et al. (1971, p. 9) that the process of measurement can be seen as the process of constructing homomorphisms, or scales, from empirical to numerical relational structures.

Not every empirical relational structure allows the construction of a homomorphism. As a result, it is useful to ask what conditions must be satisfied by the empirical relations $\succ$ and $\circ$ on A in order for it to be possible to construct a function $\phi(\cdot)$ with the desired properties. The question is important: "A measurement procedure certainly is not adequately understood if it depends on properties that are not explicitly recognized" (Krantz et al., 1971, p. 6). The question can be approached formally, by asking what axioms are necessary and sufficient for it to be possible to construct a function $\phi(\cdot)$ with the desired properties (Krantz et al., 1971, p. 8). The answer is provided 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, p. 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,

---

[4] If a homomorphism $\phi(\cdot)$ is one-one, it is said to be an *isomorphism*. In the context of our example, this will occur if no two rods are equal with respect to length, so that each rod gets assigned a unique number.

and so on, Krantz notes, it may seem as if they "introduce ontological presuppositions" and presuppose the existence of unobservable entities (Krantz, 1991, p. 3). However, he argues, measurement theorists do not need to take a position on the issue of the ontological status of such entities; the only thing measurement theorists need to assume is the existence of an empirical relational structure, that is, a set of observable entities and an observable ordering satisfying certain conditions (Krantz, 1991, p. 3). In order to justify talk about length, then, the only thing that needs to be assumed is the existence of an empirical structure consisting of a set of rods and of relations satisfying the right conditions. It is important to notice, by the way, that measurement theory is not an attempt to provide operational definitions of theoretical terms in the manner of Percy W. Bridgman (1927). Krantz et al. write that to treat measures as "objective definitions of unanalyzed concepts" is a temptation that must be resisted (Krantz et al., 1971, p. 32).

In sum, the measurement-theoretic approach emphasizes concepts like observable orderings, homomorphisms, and representation theorems. Those operating in accordance with the measurement-theoretic approach take as their starting point data concerning observable choices. The evidence employed consists of claims to the effect that a given relation satisfies certain axioms, which can be established empirically by applying a series of observable operations. From such evidence, those who use this approach reason deductively: they offer mathematical proofs or other forms of reasoning with the feature that if their premisses are true then the conclusion must be true too. It is sometimes useful to think of the defense of a given measure as a matter of constructing an argument to the conclusion that the measure represents that which it purports to represent. If so, the measurement-theoretic approach requires a

deductive argument from the premiss that some observable ordering satisfies certain conditions to the conclusion that the measure represents that which it purports to represent.

2.2 The psychometric approach

The psychometric approach is due in large part to the American Psychological Association's "Technical recommendations for psychological tests and diagnostic techniques" (1954) but was further developed by Lee J. Cronbach and Paul E. Meehl (1955). In order to develop a better idea of the nature of the psychometric approach, I will rely on two standard textbooks: *Psychometric Theory*, by Jum C. Nunnally and Ira H. Bernstein (1994), and *The New Psychometrics*, by Paul Kline (1998). As we will see, this approach emphasizes latent constructs, reliability and construct validity.

A central term of the psychometric approach is that of *reliability*. To say that a measure is reliable 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, p. 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, p. 29). A test has *internal consistency* insofar as "each item [of the test administered] measures the same variable" (Kline, 1998, p. 30). If total test scores vary too much over time, or individual items of the test diverge too much from each other, the measures are considered 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, p. 34; cf. Nunnally & Bernstein, 1994, p. 83). Again, there are several types of validity, including *face*

validity, concurrent validity, predictive validity, content validity, and *discriminant validity* (Kline, 1998, pp. 34-37). Here I will focus on *construct validity,* which is sometimes described as subsuming the other kinds (Judd & McClelland, 1998, p. 203), and which at any rate is the form of validity most relevant for present purposes. Introducing the topic of construct validity, Nunnally and Bernstein (1994, p. 84) argue that the aim of science – including social and behavioral science – is to establish functional relationships between variables of different kinds. 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 & Bernstein, 1994, p. 85).

In this spirit, two of scientists' most important tasks are to develop measures of individual constructs and to establish functional relationships between such measures (Nunnally & Bernstein, 1994, p. 85). The process of construct validation serves to help scientists reach these goals.

In practical terms, construct validation – which is often described as an instance of ordinary hypothesis testing (Cronbach & Meehl, 1955, p. 300; Johnson, 2001, p. 11316) – has three phases:

> (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 & Bernstein, 1994, pp. 86-87).

In what follows, I will discuss the three phases in order. First, scientists need to identify a class of observable variables that are potentially related to the construct. Nunnally and Bernstein argue that there is no algorithm that can be followed in this phase; instead, scientists must rely on intuition and preconceived ideas about how the construct relates to those variables. The identification of such a class of variables is a necessary condition for the second phase, which is to explore whether the different measures can be described as representing the same thing or not. This second phase is performed by "determining how well the measures of observables 'go together' (intercorrelate) empirically" (Nunnally & Bernstein, 1994, p. 88). Typically, this amounts to computing a range of correlation coefficients and examining the resulting patterns of variances and covariances. Nunnally and Bernstein 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 & Bernstein, 1994, p. 90).

The third and final phase is to show that a set of highly correlated observables in a domain can legitimately be thought of as 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 & Bernstein, 1994, p. 90). To see whether a set of intercorrelated variables can be assumed to represent a given construct like anxiety, stress or happiness, the scientist needs to explore whether the variables vary across conditions approximately like we would expect

degrees of anxiety, stress or happiness to do. In all, Nunnally and Bernstein propose that construct validity obtains if "the supposed measure(s) of the construct *behave as expected"* (Nunnally & Bernstein, 1994, p. 90).

In sum, the psychometric approach emphasizes concepts like latent constructs, reliability, and construct validity. Those operating in accordance with the psychometric approach take as their starting point data concerning measures of latent constructs and/or observable variables. The evidence employed consists of claims to the effect that the measures behave as expected, which can be established empirically by showing that the pattern of variances and covariances exhibited by the measures conform to expectations. This approach, then, does not require the existence of observable orderings satisfying conditions like transitivity; it requires a set of assumptions about a latent, unobservable construct. From such evidence, those who use this approach reason inductively: they offer reasons that give good grounds for thinking that the measure is valid. If we think of the defense of a given measure as a matter of constructing an argument, the psychometric approach requires an inductive argument from the premiss that some measure behaves as expected to the conclusion that the measure represents that which it purports to represent.

2.3 Discussion

In this section, I have outlined the two approaches to measurement in social and behavioral science: the measurement-theoretic approach and the psychometric approach. We have seen that there are real differences between the two approaches. While those operating in accordance with the measurement-theoretic approach reason deductively from the claim that a given empirical relation satisfies certain axioms, those operating in accordance with the psychometric approach reason inductively from the claim that a given measure behaves as expected. Hence, when

showing that a given measure represents that which it purports to represent – that is, when showing that a given measure is valid – the two approaches differ both when it comes to the character of the data, the nature of the evidence, and the mode of inference.

The differences between the two approaches have deep implications for the nature of measurement in social and behavioral science. Since it is possible to satisfy the strictures imposed by the one approach to measurement without satisfying those imposed by the other, a measure that has been validated in accordance with the one approach has not necessarily been validated in accordance with the other. Moreover, the simultaneous endorsement of the two approaches to measurement would lead to inconsistency, since the measurement-theoretic approach entails that an observable ordering satisfying certain axioms is necessary for measurement whereas the psychometric approach entails that it is not. Like in the case of the underlying conceptions of well-being, a failure to notice the differences between the two approaches can generate a great deal of confusion; the confusion may concern the data, evidence, and mode of inference required to establish the validity of a given measure.

Some advantages and disadvantages associated with the two approaches are readily apparent. The fact that the measurement-theoretic approach relies on deductive reasoning is *ceteris paribus* an enormous advantage. A valid deductive argument is by definition *truth-preserving*, in the sense that it cannot lead from true premisses to a false conclusion, and *erosion-proof*, in the sense that it cannot be made invalid by the introduction of new premisses; an inductive argument is neither truth-preserving nor erosion-proof (Salmon, 1992, p. 11). The Achilles heel of the measurement-theoretic approach, as we will see, is the fact that it requires observable orderings that satisfy strict axioms. The fact that the psychometric approach does not require the existence of such orderings, and can proceed from weaker premisses, is its central

advantage. The advantage can be ascribed to the fact that it relies on inductive reasoning. An inductive argument can proceed from weaker premisses because it is *ampliative*, in the sense that its conclusion has informational content that goes beyond that of its premises; deductive reasoning is not (Salmon, 1992, p. 11).

3 Approaches to welfare measurement

As we saw in the introduction, orthodox economic welfare measures differ from subjective measures of well-being with respect to the underlying conception of well-being. But there are other differences as well. In this section, I will maintain that orthodox economic welfare measures – or *economic measures*, for short – and subjective measures of well-being are based on radically different approaches to measurement: whereas economic measures are based on the measurement-theoretic approach, subjective measures are based on the psychometric approach. As we will see below, the fact that different measures are based on different approaches to measurement has deep implications for the measurement of welfare.

3.1 Economic welfare measures

Orthodox economists' arsenal of welfare measures contains a variety of measures. Pigou himself favored the *national dividend* – that is, "that part of the objective income of the community, including, of course, income derived from abroad, which can be measured in money" (Pigou, 1952, p. 31). Measures of this general kind, like *gross domestic product* (GDP) per capita, have well-known shortcomings but continue to be widely used for public policy purposes (Nussbaum & Sen, 1993, p. 2). The importance of the national product as a measure of well-being helps explain the widespread concern with economic growth: since "growth" is often used to refer to

the first derivative of the national product, and "growth rate" to refer to the second derivative, high growth (or growth rate) is thought to be an indication of future well-being.

An alternative way to evaluate the welfare consequences of policy interventions is in terms of *consumer surplus* (CS) and *producer surplus* (PS). The notion of consumer surplus goes back to Jules Dupuit, who in a paper originally published in 1844 wished to determine the conditions under which public works – such as the building of a bridge – can "be declared of public utility" (Dupuit, 1969, p. 255). Dupuit's ideas were developed and popularized by Alfred Marshall (1920; 1st ed. 1890), who defined the consumer surplus of a good as "[the] excess of the price which [the consumer] would be willing to pay rather than go without the thing, over that which he actually does pay" (Marshall, 1920, p. 124). Measures of consumer and producer surplus are widely used to evaluate the consequences of public policy: "Consumer surplus is the overwhelming choice as a welfare indicator" (Slesnick, 1998, p. 2110). Moreover, consumer/producer surplus is the tool preferred by many economics textbooks when evaluating the welfare consequences of interventions like price ceilings and trade restrictions (cf. Mankiw, 2001, "Part III: Markets and welfare").

Yet another set of measures revolve around the concepts of *compensating variation* (CV) and *equivalent variation* (EV). These notions were developed in a series of publications by John R. Hicks (e.g., 1943), who had noted certain technical difficulties associated with surplus measures. The CV is "the amount of money which, when taken away from an individual after an economic change, leaves the person just as well off as before," while the EV is "the amount of money paid to an individual which, if an economic change does not happen, leaves the individual just as well off as if the change had occurred" (Just, Hueth, & Schmitz, 2004, p. 9). CV/EV measures are used in many contexts to assess changes in welfare:

> In cost-benefit analysis and other exercises in applied welfare economics, *aggregate willingness-to-pay* – the simple sum of Hicksian compensating variations, is often used as a test. A positive sum is taken as evidence of a social improvement or an increase in economic efficiency (Blackorby & Donaldson, 1990, pp. 471-472).

Though different, all these measures have much in common, including the fact that they are all based on the measurement-theoretic approach to measurement.

There is much evidence that orthodox economists operate with the measurement-theoretic approach. First, there is a purely historical connection, in that both orthodox economics and the measurement-theoretic approach were shaped by the same broadly speaking empiricist philosophy of science. Although economists during the early twentieth century were comfortable talking about mental states like pain and pleasure – Pigou (1952, p. 10), as a case in point, defined welfare in terms of "states of consciousness and, perhaps, their relations" – economists soon, under the influence of logical empiricism in philosophy, behaviorism in psychology, and operationalism in physics, came to see references to unobservable mental states as unscientific and at any rate dispensable (Angner & Loewenstein, 2008). Much for the same reason, the measurement-theoretic approach was explicitly designed in order to provide a solid foundation for measurement on the basis of observable relations such as "*x* is longer than *y*" while avoiding controversial "ontological presuppositions." As J. D. Trout points out, this approach "has been developed in a way that displays the distinct and deep influence of empiricism" (Trout, 1998, p. 49). Moreover, as we will see shortly, the measurement-theoretic approach was in fact motivated in part by the problem of utility measurement (Krantz et al., 1971, p. 9).

Second, the manner in which economic measures are defended is perfectly consistent with the measurement-theoretic approach. In order to see how utility measurement is addressed within the measurement-theoretic framework, let us return to Krantz (1991), who discusses the issue explicitly. 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 commodity bundles, and R is a binary relation such that aRb means that a is chosen over b in a pair-wise choice in which both a and b are available. Krantz 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, p. 28).[5]

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, p. 28). This view, of course, also assumes that choices reflect preferences over the various options. Anyway, as Krantz notes, "most utility theories cling to the idea that the ordering is based on observation of choice behavior" (Krantz, 1991, pp. 28-29).

In order for the representation theorem to work, the measurement theorist must assume 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) note, transitivity is a *necessary* condition, in the sense that it is mathematically necessary for any representation theorem to work. Just like in the case of the measurement of length, the axioms are seen as empirical, descriptive laws; in this case, they are laws of choice. Thus, for instance, the famous axioms articulated by von Neumann and Morgenstern (1944) "constitute a set of qualitative laws for

---

[5] The second point has to do with whether utility scales are ordinal or cardinal, which is a topic I will not go into here.

'rational' decisions among risky options" (Krantz, 1972, p. 1428). Very often, of course, these axioms are treated as normative principles. In the present context, however, the normative status of the axioms is irrelevant: the point is that the representation theorem requires that they be true descriptive laws. For the theorem to apply, the choice structure must *in fact* satisfy the axioms, that is, the axioms must be *true* of the relevant empirical relational choice structure. As Norman Cliff puts the point: "Axiomatic measurement theory holds that … the order relations … must display a strong kind of consistency" (Cliff, 1992, p. 187).

We can now see how the measurement of utility is supposed to relate to the example of length measurement. 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. Just like in the case of length, measurement theory is supposed to allow us to remain agnostic about the existence of unobservable attributes; utility is understood as a measure, or *index*, of preference satisfaction, with no connection whatsoever with mental states or other unobservables. Thus, the measurement theorist claims to provide methodological foundations for talk about utility and preference in a manner that is ontologically non-committal, without, as the case is sometimes made, "pretending to look into the head of the agent." In order to allow the construction of a representation theorem, we need to identify a set of axioms, which are seen as empirical, descriptive laws of choice. One of these laws is the notion that preferences are transitive. What 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, again, the importance of the consistency or transitivity condition: if this condition is not in fact satisfied, measurement theory offers no grounds for constructing a utility function on the set of available options.

When orthodox economists defend their welfare measures, their actual practice conforms perfectly to this picture. Those economists begin with the assumption that market choices satisfy certain axioms, and on the basis of the axioms, proceed to offer a formal proof that the measure is an index of preference, i.e., a utility function. This is true for all three kinds of welfare measure discussed above. The proofs are available in any standard-issue graduate-level microeconomics textbook like Andreu Mas-Colell, Michael D. Whinston, and Jerry R. Green's *Microeconomic Theory* (Mas-Colell, Whinston, & Green, 1995). Mas-Colell et al. show that given a number of assumptions, e.g., about the rationality of individuals and the nature of the budget set, and holding prices fixed, it can be shown that utility (understood in terms of preference satisfaction) is strictly increasing in individual wealth, which is to say that under certain assumptions wealth is a utility function (Mas-Colell et al., 1995, p. 56). Given a similar set of assumptions, they also show that consumer surplus as well as compensating and equivalent variation are utility functions (Mas-Colell et al., 1995, pp. 81-83). Because this procedure establishes that the measure is a homomorphism, the actual practice of orthodox economists indicates a commitment to the measurement-theoretic approach to measurement.

Finally, orthodox economists show few signs of a commitment to the psychometric approach. They typically do not even mention latent constructs, reliability, or construct validity. Moreover, they make little to no effort to validate their measures in the manner required by the psychometric approach, that is, by showing that a given measure exhibits the appropriate pattern of variances and covariances. Certainly, Mas-Colell et al (1995) waste no time or ink on such things. The reason why orthodox economists came to think of CS/PS measures as inferior to CV/EV measures, for example, is not that the former were found to have a lower validity coefficient than the latter in psychometric studies, but that in the presence of wealth effects

standard consumer surplus measures may fail to be utility functions, that is, homomorphisms (cf. Mas-Colell et al., 1995, pp. 88-91).

3.2 Subjective measures of well-being

What I have called "subjective measures" in fact includes a range of specific measures. For most of their history, subjective measures have been constructed on the basis of one or more straightforward questions like that of Gurin et al. (1960), quoted in the introduction. Sonja Lyubomirsky and Heidi S. Lepper (1999) offer four prompts of the form "In general, I consider myself…" and invite subjects to respond on a seven-point scale, where 1 represents "… not a very happy person" and 7 "… a very happy person" (Lyubomirsky & Lepper, 1999, p. 151). Others ask subjects "How do you feel about your life as a whole?" and give them response categories ranging from "Delighted," "Pleased," and "Mostly satisfied," through "Mixed (about equally satisfied and dissatisfied)" to "Mostly dissatisfied," "Unhappy" and "Terrible" (Andrews & Withey, 1976, p. 18). Occasionally, researchers invite responses using graphic representations like horizontal lines (Watson, 1930), ladders and mountains (Cantril, 1965), or happy and sad faces (Andrews & Withey, 1976). Some of these questions were designed to represent affective states, some to represent cognitive states, and some to represent a combination of the two (Angner, 2010b).

A somewhat different approach has been developed by Daniel Kahneman (1999) and others under the heading of *experience sampling.* Kahneman prompts his subjects every so often – e.g. with the use of handheld electronic devices – to judge the "quality of their momentary experience" along the "good/bad dimension" (Kahneman, 1999, p. 7). The assumption is that, at every point in time, the brain rates the quality of experience in a manner that can be represented on a single numerical scale and which, furthermore, is accessible to the agent. What matters, at

the end of the day, is the time integral (which Kahneman calls *objective happiness*) of the instant happiness rating (which Kahneman calls *subjective happiness*) (Kahneman, 1999, p. 5). The effort to produce a dense record of an individual's affective state as a function of time was pioneered by Hornell Hart (1940), the inventor of the *euphorimeter:* a device that would permit the quick assessment of an individual's level of happiness based on self-reports. Though Kahneman and co-authors have since developed other measures, they insist: "Experience sampling is the gold standard" (Kahneman et al., 2004b, p. 1777).

More recently, Kahneman and Alan B. Krueger have suggested the use of a measure they call the *U-index* (Kahneman & Krueger, 2006; cf. Krueger, 2009). Introduced under the heading of "A Measure of Society's Well-Being," the U-index is clearly intended to be a measure of social well-being. The "U" stands for "unpleasant" or "undesirable," and the index "measures the proportion of time an individual spends in an unpleasant state," where an episode gets classified as pleasant or unpleasant depending on whether the strongest affect experienced during the episode is positive or negative (Kahneman & Krueger, 2006, pp. 18-19). The U-index was designed to overcome several perceived problems associated with other subjective measures, above all problems related to interpersonal comparability (Krueger, 2009, p. 3).

There is much evidence that proponents of subjective measures operate with the psychometric approach. First, there is a purely historical connection, in that both subjective measures and psychometrics grew out of personality psychology, which uses tests based on self-reports in order to assess individual differences in personality, and which emerged shortly after World War I. As personality psychologists grew confident that they could explore scientifically various personality characteristics, it was not long before they started asking questions about the determinants and distribution of positive or desirable mental states like happiness and

satisfaction. And psychometrics too was a direct outgrowth of personality psychology: from its very beginnings, personality psychology was characterized by an emphasis on measurement and psychometrics, and by a desire to be useful to corporations and governments (Winter & Barenbaum, 1999, p. 5).

Second, the manner in which subjective measures are defended is perfectly consistent with the psychometric approach. In order to show this, I will rely on a 1999 review article by Ed Diener, Eunkook M. Suh, Richard E. Lucas, and Heidi L. Smith (1999). In defense of subjective measures, the authors write:

> 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, pp. 277-278).[6]

The reasoning invoked here fits the schematic picture painted in section 2.2 very well. First, Diener et al. identify a set of variables – in this case, spouses' reports, the absence of suicides, and so on – that they take to reflect the same construct as self-reported happiness. Second, the authors explore the degree to which these variables intercorrelate (positively and negatively) with the self-reports. Third, because all these variables are found to correlate positively, Diener et al. conclude that they all represent the same construct. Moreover, Diener et al. check whether the construct behaves 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 as anticipated.

---

[6] References have been omitted.

A more recent contribution makes the point even clearer. Diener, Richard Lucas, Ulrich Schimmack, and John Helliwell (2009) dedicate a whole chapter to the thesis that "Well-Being Measures Are Valid" (Diener et al., 2009, p. 67). The authors argue that the general procedures that must be followed when validating a scientific measure are straightforward and uncontroversial; that widely used happiness measures have passed the basic tests mandated by these procedures; and that those who would reject the validity of happiness measures have failed to marshal any empirical evidence against them. Specifically, Diener et al. point out that "[a] key requirement for any measure is that it is reliable," and proceed to outline the evidence for the reliability of measures of subjective well-being (Diener et al., 2009, pp. 68-74). After that, the authors introduce the concept of validity – in the process describing face validity, content validity, convergent validity, and discriminant validity – and the process of construct validation, and they proceed to argue that subjective measures are valid in the relevant sense (Diener et al., 2009, pp. 74-93).

In fact, efforts to validate measures of happiness in accordance with what we now call the psychometric approach are as old as the measures themselves. Throughout the history of these measures, psychologists have postulated the existence of a construct like happiness, satisfaction, or similar; proposed a measure of it; and sought to confirm that the measure behaves as expected when compared to objective life circumstances, other people's judgment of subjects' happiness, measures of mental health, and so on (cf. Lyubomirsky & Lepper, 1999, p. 145). In an early review of the literature, Warner Wilson made the case by arguing that self-reported happiness scores were sufficiently correlated with the judgments of associates, teachers, professors, principals, psychologists, and clinical judges, as well as with scores on elation-depression scales (Wilson, 1967, pp. 294-295). Hart found that happiness scores, tracked over time, changed as

expected when participants fell in love, experienced the death of their mothers, or contemplated suicide (Hart, 1940, pp. 19-25). Based on a 1997 review, subjective measures are sufficiently positively correlated with happiness ratings of friends and family, psychologists' judgments, amount of smiling; sufficiently negatively correlated with depression; and not overly correlated with general intelligence, current mood, humility, and the language in which the question was asked (Diener & Suh, 1997, pp. 436-438).

Because the procedure followed by all these authors serves to establish the reliability and validity of measures of a latent construct, the manner in which proponents of subjective measures defend their measures exhibits the hallmarks of the psychometric approach. The contention that proponents of subjective measures operate with the psychometric approach to measurement is further supported by the fact that they explicitly refer to "psychometric criteria" (Lyubomirsky & Lepper, 1999, p. 140) and "psychometric properties" (Diener et al., 1999, p. 277). In brief, the actual practice of proponents of subjective measures indicates a commitment to the psychometric approach to measurement.

Finally, proponents of subjective measures show few signs of a commitment to the measurement-theoretic approach. They typically do not even mention observable orderings, homomorphisms, or representation theorems, and they make little to no effort to defend their measures in the manner favored by the measurement-theoretic approach, that is, by identifying an empirical relational structure and proving that their measures are homomorphisms.[7] Speaking about the measurement of attitudes – conceived of as cognitive, affective, or conative mental

[7] There are contributions to the literature on subjective measures that aspire to use an axiomatic approach, which makes them in this respect consistent the measurement-theoretic approach (Kahneman, Wakker, & Sarin, 1997). I thank Anna Alexandrova for pointing this out.

states – Dawes and Smith note: "Representational measurement is rare in the field of attitude; instead this field is permeated by questionnaires and rating scales" (Dawes & T. L. Smith, 1985, pp. 511-512). Indeed, for reasons we will explore in section 4, within contemporary psychology, the measurement-theoretic approach is widely regarded as a failure and is not commonly used (Cliff, 1992, p. 189; John & Benet-Martínez, 2000, p. 341).

3.3 Why are economic and subjective measures so different?

So far, I have argued that orthodox economic welfare measures and subjective measures of well-being are based on radically different approaches to measurement: whereas economic measures are based on the measurement-theoretic approach, subjective measures are based on the psychometric approach. The thesis that economic and subjective measures are based on different approaches to measurement helps explain the fact that proponents of different kinds of measure go about measuring well-being so differently, and consequently end up with such different measures. The remarkable explanatory power of the thesis, by the way, constitutes additional evidence in its support.

First, the thesis helps account for the fact that proponents of different measures rely on such different data. As we have seen, proponents of subjective measures are comfortable using questionnaire and survey data, whereas orthodox economists tend to require data concerning observable choices. The contention that proponents of subjective measures adopt the psychometric approach helps explain why it comes so naturally for them to take questionnaire and survey data as their starting point: questionnaires and surveys have been part and parcel of the practice of psychometrics since the very beginning. The contention also explains why proponents of subjective measures admit self-reports: it is not that they unquestioningly believe that people know how happy they are, but rather that self-reports can constitute one of many

measures of the latent construct of interest. Meanwhile, the contention that orthodox economists adopt the measurement-theoretic approach helps explain why it comes so naturally for them to take data about economic transactions like market choices as their starting point: since the very beginning, the measurement-theoretic approach has required observable orderings like those imposed on a set of alternatives by an agent's choices. Orthodox economists' preference for observable choices is well documented. As Amartya Sen notes: "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, p. 71). The reasoning underlying the exclusive reliance on choice data is neatly expressed in *The Welfare Economics of Public Policy* (Just et al., 2004). After pointing out that utility is not observable, the authors 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., 2004, p. 98).

Second, the thesis helps account for the fact that proponents of different measures rely on such different evidence. As we have seen, proponents of subjective measures rely on evidence to the effect that subjective measures behave as expected, which can be established empirically by showing that the pattern of variances and covariances exhibited by the measures conform to expectations, while orthodox economists rely on evidence to the effect that a market-based choices satisfy certain axioms, which can be established empirically by applying a series of observable operations. The contention that proponents of subjective measures adopt the psychometric approach explains why it comes so naturally for them to use the kind of evidence they do: the comparison of patterns of variances and covariances with expectations forms the

very heart of the approach. At the same time, the contention that orthodox economists adopt the measurement-theoretic approach explains why it comes so naturally for them to use the kind of evidence that they do: claims to the effect that certain orderings satisfy particular axioms are a critical part of the measurement-theoretic approach.

Third, the thesis that economic and subjective measures are based on different approaches to measurement helps account for the fact that proponents of different measures use such different arguments in support of their measures. As we have seen, proponents of subjective measures validate their measures by establishing reliability, construct validity, and so on, whereas orthodox economists prove that a particular function is a utility function. The contention that proponents of subjective measures operate with the psychometric approach helps explain why they believe that the use of a given measure has been justified when it has been shown that it behaves as expected when compared to measures of other latent constructs, observable behavior, and so on: because this procedure, if successful, establishes reliability and construct validity, which is what the psychometric approach requires. Meanwhile, the fact that orthodox economists operate with the measurement-theoretic approach helps explain why they 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 a utility function: because this procedure, if successful, in fact establishes that the measure is a homomorphism, which is what the measurement-theoretic approach requires.

Fourth, the thesis explains why orthodox economists' approach to welfare measurement (including their choice of data, evidence, and mode of inference) strikes proponents of subjective measures as terribly inadequate, and *vice versa*. For anyone trained in the psychometric tradition, a person's income and market choices are likely to seem hopelessly indirect as measures of her

welfare, independently of whether welfare means preference-satisfaction, happiness, or satisfaction. Those trained in psychometrics have been taught to reject measures not validated in the manner of the psychometric approach, and as we have seen, orthodox economists make no effort to validate their efforts in this manner. For somebody trained in the measurement-theoretic tradition, meanwhile, a person's answers to questionnaires and surveys are likely to appear inherently dubious. Those trained in this tradition have been taught to reject measures that have not been shown to be homomorphisms in the manner of the measurement-theoretic approach, and as we have seen, proponents of subjective measures make no effort to establish that their measures are homomorphisms based on observable orderings. Moreover, the sort of inductive arguments required by the process of construct validation are likely to appear far inferior to the deductive reasoning required by the measurement-theoretic tradition.

Finally, the thesis helps account for the fact that orthodox economists, on the one hand, and psychologists and unorthodox economists, on the other, adopt such different accounts of well-being. As we saw in the introduction, whereas orthodox economists tend to adopt preference-satisfaction accounts of well-being, proponents of subjective measures tend to adopt mental-state accounts. Orthodox economists were originally drawn to preference-satisfaction accounts after convincing themselves that pleasure and pain could not, whereas preference satisfaction could, be measured scientifically (Angner & Loewenstein, 2008, p. ???; Mandler, 1999, p. 6). Similarly, twentieth-century interest in happiness appears, at least in part, to be a result of traditional psychologists and unorthodox economists convincing themselves that mental states like happiness can, in fact, be scientifically measured. An uncharitable interpretation of these events suggests that economists and psychologists alike are like the drunken man who looks for his lost key not where he lost it, but where the light is. A more charitable interpretation

of these events is that psychologists and economists alike believe that accounts or conceptions of well-being can be judged in part on the grounds of whether they permit the development of adequate measures of well-being. 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, p. 1). 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, p. 54).

If economists and psychologists alike believe that accounts of well-being can be judged in part on the grounds of whether they permit the development of adequate measures of well-being, this would go a long way toward explaining why orthodox economists tend to favor preference-satisfaction accounts whereas psychologists and unorthodox economists tend to favor mental-state accounts.[8] According to this hypothesis, the fact that orthodox economists operate with the measurement-theoretic approach to measurement helps explain why they reject mental-state accounts of well-being on favor of preference-satisfaction accounts. After all, their favored approach requires that measurement be based on observable orderings, but because they believe that no such ordering exists in the case of happiness measurement, they believe that happiness cannot be measured, which they take to be reason to reject mental-state accounts of well-being;

[8] Below, we will see that there are exceptions to this rule, since there are economists who appear to gravitate toward the psychometric approach to measurement while remaining committed to preference-satisfaction accounts of well-being.

at the same time, they believe that degrees of preference satisfaction can be measured based on an observable ordering, so they believe that degrees of preference satisfaction can be measured, which they take to be a reason to adopt some preference-satisfaction accounts of well-being. According to my hypothesis, meanwhile, the fact that psychologists and unorthodox economists operate with the psychometric approach helps explain why they, by and large, adopt mental-state accounts of well-being. After all, their favored approach requires measures that behave as expected rather than observable orderings; because they believe that existing measures of happiness, satisfaction, etc., do behave as expected, they believe that happiness, satisfaction, etc., can be measured, which they take to be a reason to adopt some mental-state accounts of well-being. In sum: the difference between the two approaches to measurement helps explain why subjective measures are based on mental-state accounts, whereas orthodox economic measures are based on preference-satisfaction accounts of well-being.

3.4 Discussion

In this section, I have defended the thesis that economic measures are based on the measurement-theoretic approach while subjective measures are based on the psychometric approach. In addition, I have maintained that this thesis has remarkable explanatory power. The thesis that subjective and economic measures are based on different approaches to measurement serves to bring out radical differences between the two kinds of measure, *inter alia*, when it comes to the nature of the data, the character of the evidence, and the mode of inference. The thesis also helps explain orthodox economists' rejection of subjective measures, and *vice versa*, as well as why those who operate with the psychometric approach tend to favor mental-state accounts whereas those who operate with the measurement-theoretic approach tend to favor preference-satisfaction

accounts. (As we will see below, however, the measurement-theoretic approach to measurement is logically consistent with a commitment to preference-satisfaction accounts of well-being.)

Notice that my thesis does not presuppose that those who are committed to the measurement-theoretic approach never use inductive reasoning or that those who are committed to the psychometric approach never use deductive reasoning. Even for somebody fully committed to the measurement-theoretic approach, inductive reasoning might be necessary, e.g., when trying to assess a person's preferences over an infinitely large set of alternatives (perhaps represented by a vector of real numbers) on the basis of a necessarily finite series of observable choices; moreover, deductive reasoning is widely used in all the sciences. Similarly, my thesis does not presuppose that those who rely on the psychometric approach never use observable-choice evidence: the psychometric approach permits, though it does not require, the use of such evidence. Hence, my thesis is not undercut by the observation that orthodox economists sometimes rely on inductive reasoning, that psychometricians sometimes use observable-choice evidence, and so on.

Attention to the differences between the two kinds of measure helps illuminate some of their advantages and disadvantages. The fact that economic measures are defended using deductive arguments is *ceteris paribus* a huge advantage, since (as pointed out in section 2.3) such arguments are both truth-preserving and erosion-proof. At the same time, however, the fact that these arguments proceed from the assumption that observable choices constitute an ordering satisfying strict axioms is a disadvantage, since that is a strong assumption indeed. The fact that subjective measures are defended using inductive arguments is *ceteris paribus* a disadvantage, since (as we also saw in section 2.3) such arguments are neither truth-preserving nor erosion-proof. Yet, because inductive reasoning is ampliative, these arguments do not require observable

orderings and can proceed from the weaker assumption that certain measures behave as expected.

4 The shift toward to the psychometric approach

Given the thesis that economic measures are based on the measurement-theoretic approach to measurement while subjective measures are based on the psychometric approach, it should not be surprising that the rise of subjective measures would be associated with a transition from the measurement-theoretic to the psychometric approach. In this section, I argue that this trend is evident also outside of the literature on subjective measures, among authors who apparently remain committed to preference-satisfaction accounts of well-being. Though this trend has gone largely unnoticed, there are exceptions. In his contribution to the previously mentioned volume about the U-index (Krueger, 2009), William Nordhaus remarks that the strategy of Krueger and co-authors "uses a completely different approach to measuring the values associated with time uses – one based on surveys or other psychometric measurements" (Nordhaus, 2009, p. 127). Nordhaus's use of the term "psychometric measurement" strongly suggests that he drawing attention to a trend toward the psychometric approach to measurement.

The transition from the measurement-theoretic to the psychometric approach to measurement is most evident among economists who jettison orthodox economic welfare measures in favor of subjective measures of well-being. These economists exhibit a marked shift from observable choice data toward questionnaire and survey data; from evidence showing that some ordering satisfies strict axioms toward evidence showing that a given measure behaves as expected; from mathematical proofs and other forms of (truth-preserving, erosion-proof) deductive arguments toward (ampliative) inductive arguments; and from preference-satisfaction

accounts of well-being toward mental-state accounts. As pointed out in section 3.4, the psychometric approach can make use of, though it does not require, measures of observable choices, deductive reasoning is widely used in all the sciences, and so on. My thesis offers a plausible account of the rise of subjective measures of well-being and the return of mental-state accounts of well-being, as both of these developments can be seen as resulting from a shift from the measurement-theoretic to the psychometric approach to measurement, in combination with mounting evidence of reliability and validity of subjective measures.

The trend from the measurement-theoretic toward the psychometric approach is evident also outside of the literature on subjective measures, among authors who apparently remain committed to preference-satisfaction accounts of well-being. In this sense, I will argue, current trends in economics mimic recent trends in psychology. In contemporary psychology (as mentioned in section 3.2 above), the measurement-theoretic approach is largely regarded as a failure and not widely used. The main reason is that in the social and behavioral sciences observable orderings satisfying conditions like transitivity are rare. As Cliff puts the problem: "Measurement theory says that if certain conditions hold, then scales of a given kind are defined. If not, they are not. But data always contradict one or the other axiom" (Cliff, 1992, p. 189). By 1991, Krantz himself had come to a very similar conclusion. Under the heading "The Myth of Utility," Krantz argues:

> Choice does indeed depend on the method of testing and depends especially on how options are *framed.* Results such as these show that ordering options by choices is no more determinate than ordering "overall" reading skill by testing on a particular set of materials (Krantz, 1991, p. 32).

Krantz points to empirical results by Tversky and Kahneman and their co-authors, whose work is widely interpreted as showing that real-life choices reflect what is often called "normatively irrelevant" factors and consequently fail to satisfy the axioms. Krantz infers: "Preference

ordering is a behaviorist myth" (Krantz, 1991, p. 35). From our vantage point, if anything, the case for the empirical adequacy of the axioms of rational choice theory is even weaker than it was when Krantz wrote his retrospective two decades ago. Many different researchers claim to have uncovered 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 (Camerer & Loewenstein, 2004; Kahneman, 2003; Rabin, 2002).

While realizing that the measurement-theoretic approach is largely inapplicable due to the lack of relevant observable orderings, psychologists have become increasingly convinced of the adequacy of the psychometric approach. As Charles M. Judd and Gary H. McClelland put it:

> While the revolution of representative measurement was sputtering, psychometric measurement was faring better in its ability to make predictions and was making progress at getting its own house in order with respect to being able to test its measurement scales (Judd & McClelland, 1998, p. 183).

Hence, contemporary psychologists have come to believe not just that the measurement-theoretic approach is inadequate, but that the psychometric approach has much to be said for it in terms of building theories with predictive ability. Consequently, the psychometric approach has come to dominate the field: "Most measurement in social psychology consists of questionnaire and observational measures whose validity is established not by a set of axiomatic consistency tests but rather by the observed patterns of variances and covariances that they display" (Judd & McClelland, 1998, p. 201).

My thesis in this section is that a similar trend is apparent in contemporary economics. In the past few years, a number of economists have noticed the problem posed by non-standard choice behavior for welfare measurement, and have worked to make sense of the idea of welfare measurement in the absence of observable orderings satisfying the relevant axioms. In a 2007 working paper, for example, Jerry Green and Daniel Hojman (2007) develop a method that

permits them to assess the welfare of a decision maker whether or not his or her choices satisfy the axioms of rational choice theory; the idea is to assume the existence of unobservable, simultaneously-held preference relations, and use observable choices to estimate properties of those preference relations. In a similar vein, Ariel Rubinstein and Yuval Salant (2008) maintain that, given a domain of objects, economists need to distinguish what they call "mental preference" – an unobservable "mental attitude of an individual towards to objects" – from observable choice and to develop techniques to estimate the former based on the latter. In their view: "There is no escape from including mental entities, such as the way in which an individual perceives the objects and his mental preferences, in economic models" (Rubinstein & Salant, 2008, p. 117). These papers represent a growing literature on how to do welfare analysis in the presence of non-standard choice behavior.

While there is much to be said about this development, what matters most for present purposes is that this proposal shows clear evidence of a transition from the measurement-theoretic to the psychometric approach to measurement. These authors clearly reject the notion that measurement must take as its starting point an observable ordering satisfying strict axioms; instead, they take as their starting point some set of assumption about unobservable entities like "mental preference," which psychometricians would call latent constructs. Moreover, these authors reject the notion that a measure of preference satisfaction must be shown to be a homomorphism based on some observable ordering; instead, they use observable evidence to make inductive inferences about the satisfaction of unobservable mental preferences. Insofar as these proposals take as their starting point an unobservable construct like a "mental preference" and reason inductively from evidence other than observable orderings, the actual practice of these economists conforms more closely to the psychometric than to the measurement-theoretic

approach. In fact, the realization that the measurement-theoretic approach is inapplicable in the absence of observable orderings of the relevant kind constitutes an implicit acknowledgement that the psychometric approach is the only game in town.

It might be objected that the economists mentioned here do not think of their work as representing a trend from the measurement-theoretic to the psychometric approach to measurement, and that this fact undercuts the thesis that such a trend is taking place. I have no reason to doubt that few of the economists mentioned here think of their activities in these terms. In fact, I think it is fair to say that many economists are unaware of the existence of two mutually inconsistent approaches to measurement, since the difference has received vanishingly little attention in the literature. According to Judd and McClelland, "there is almost no literature in psychology where the two measurement traditions have spoken to each other" (Judd & McClelland, 1998, p. 227) and much the same thing is true in economics. Nevertheless, the fact that a given economist does not think of his or her activities in these terms does not undercut my thesis. It is perfectly possible to operate with a given approach to measurement, and to shift from one approach to another, without being aware of it. Hence, the fact (if it is one) that few of the economists mentioned here think of their work as representing a trend from the measurement-theoretic to the psychometric approach to measurement does not undercut my thesis.

5 Conclusion

Why should anyone pay attention to issues of measurement? As Judd and McClelland put it: "While the individual researcher can ignore measurement issues for the most part without consequence, the discipline as a whole suffers when theories and methods underlying measurement are undeveloped and unscrutinized" (Judd & McClelland, 1998, p. 180). While

most practicing scientists can get along just fine without thinking seriously about issues of measurement, there are times when attention to the nature of measurement is critical for understanding the nature, advantages, and disadvantage of alternative approaches. It is my hope that the present discussion has served to illustrate the point at least in the case of welfare measurement.

In this paper, I have argued that the rise of subjective measures of well-being represents two major trends in contemporary welfare economics: a shift from preference-satisfaction to mental-state accounts of well-being and a shift from the measurement-theoretic to the psychometric approach to measurement. The thesis that subjective and economic measures are based on different approaches to measurement serves to bring out radical differences between the two kinds of measure, *inter alia*, when it comes to the nature of the data, the character of the evidence, and the mode of inference. Attention to these differences helps illuminate some of the advantages and disadvantages associated with the two kinds of measure. Though I am not the only one to notice the shift from measurement-theoretic to the psychometric approach to measurement – which is evident also outside of the literature on subjective measures, among authors who apparently remain committed to preference-satisfaction approach – the fact that orthodox economic welfare measures and subjective measures of well-being are based on different approaches to measurement is seldom explicitly acknowledged. Both these trends constitute radical methodological transitions, which are likely to have a significant impact on the shape of welfare economics and on the public policy that it informs.

Like in the case of underlying conceptions of well-being, the difference matters. Since it is possible to satisfy the strictures imposed by the one approach to measurement without satisfying those imposed by the other, a measure that has been properly validated in accordance

with the one approach has not automatically been validated in accordance with the other. It is quite clear that the assessment of both orthodox economic welfare measures and subjective measures of well-being will hinge, to some extent, on the adequacy of the two approaches to measurement. Moreover, a failure to notice the difference between the two approaches can generate a great deal of confusion. For example, economists often reject subjective measures of well-being by saying that mental states like happiness simply cannot be measured. In his 1975 book *Two Cheers for the Affluent Society: A Spirited Defense of Economic Growth*, Wilfred Beckerman dismissed efforts to measure happiness by declaring that "[the] concept of happiness is one for which there can be no scientific objective measure" (Beckerman, 1975, p. 53). This objection can be understood as presupposing that observable orderings are necessary for measurement, as the measurement-theoretic approach suggests. The criticism does not acknowledge that subjective measures – whatever their flaws – are defended in the manner of the psychometric approach, not the measurement-theoretic one. Similarly, psychologists sometimes criticize economists for failing to properly validate their measures in the manner of the psychometric approach. For instance, Diener et al. (2009) attack economists for failing to properly validate GDP as a measure of welfare, as they write that "economists assumed that income is a valid indicator of well-being, but few studies have systematically tested this assumption" (Diener et al., 2009, p. 75). The objection can be understood as presupposing that measures must be validated in accordance with the psychometric approach. The critique does not acknowledge that economic measures – whatever their flaws – are defended in accordance with the measurement-theoretic approach to measurement, not the psychometric one. The fact that the difference is so rarely acknowledged might explain why the differences between them are largely ignored.

It is important to notice what my thesis does not entail. There is no clear sense in which either one of the two trends *per se* represents a shift from less to more *realism*, that is, the view that science, in its theories, aims "to provide a true description of the world" (Okasha, 2002, p. 59); from less to more *realisticness*, a cluster of arguably desirable properties of scientific theories, including plausibility and usefulness (Mäki, 1998); or from lower to higher *fidelity*, meaning "messier and yet more faithful to the realities of human life" (Alexandrova & Haybron, n.d.). The trend toward mental state accounts of well-being and the measurement-theoretic approach to measurement has no direct implications concerning the aims of science or the attributes of scientific theories at all. Moreover, I see no interesting sense in which mental-state accounts and the psychometric approach *per se* are messier and more faithful to human realities; indeed, the theory of measurement might strike an innocent observer as messy enough, given that it took the authors of *Foundations of Measurement* three volumes to spell it out in the detail required to capture the relevant phenomena. (That said, I do not deny that psychometrics and/or mental-state accounts when combined with other assumptions, e.g., about the aim of science or the nature of theories, might have implications for realism, realisticness, or fidelity.)

The notion that well-being should serve as a goal – perhaps even the only goal – for public policy is hardly uncontroversial. After discussing the roles that the concept of well-being is frequently assumed to play, as quoted in the introduction, Scanlon goes on to argue that "it is a mistake to think that there is a single notion of well-being that plays all the roles I have mentioned" (Scanlon, 2000, p. 109). Yet, insofar as the concept of well-being plays any role at all in our political and other deliberations, there will be a need for a systematic way to assign numbers or other mathematical objects to individuals and groups, that is, there will be a need for measures of individual and social well-being. And if so, questions about the nature of

measurement – e.g., what it takes to be able to say that a given measure is a valid measure of well-being – will sooner or later have to be addressed. If it is indeed true that there is a strong trend away from the measurement-theoretic approach and toward the psychometric approach in contemporary economics, this would constitute a radical change in the methodological foundations of modern economics, which is likely to have a significant impact on the shape of welfare economics and on the public policy that it informs. In addition, if the methodological sophistication of psychometricians could be integrated with the statistical sophistication of econometricians, these trends could offer promising avenues of research indeed.

REFERENCES

Adler, M., & Posner, E. (2008). Happiness research and cost-benefit analysis. *The Journal of Legal Studies*, *37*(S2), S253-S292. doi:10.1086/590188

Alexandrova, A., & Haybron, D. M. (n.d.). High fidelity economics. In J. B. Davis & D. W. Hands (Eds.), *The Elgar Companion to Recent Economic Methodology*.

American Psychological Association. (1954). Technical recommendations for psychological tests and diagnostic techniques. *Psychological Bulletin*, *51*(2, Pt.2), 1-38. doi:10.1037/h0053479

Andrews, F. M., & Withey, S. B. (1976). *Social Indicators of Well-Being: Americans' Perceptions of Life Quality*. New York, NY: Plenum Press.

Angner, E. (2008). The philosophical foundations of subjective measures of well-being. In L. Bruni, F. Comim, & M. Pugno (Eds.), *Capabilities and Happiness* (pp. 286-298). Oxford: Oxford University Press.

Angner, E. (2009a). The politics of happiness: Subjective vs. economic measures as measures of social well-being. In L. Bortolotti (Ed.), *Philosophy and Happiness* (pp. 149-166). New York, NY: Palgrave Macmillan.

Angner, E. (2009b). Subjective measures of well-being: Philosophical perspectives. In H. Kincaid & D. Ross (Eds.), *The Oxford Handbook of Philosophy of Economics* (pp. 560-579). Oxford: Oxford University Press.

Angner, E. (2010a). Are subjective measures of well-being "direct"? *Australasian Journal of Philosophy*. doi:10.1080/00048400903401665

Angner, E. (2010b). Subjective well-being. *Journal of Socio-Economics*, *39*(3), 361-368. doi:10.1016/j.socec.2009.12.001

Angner, E., & Loewenstein, G. (2008). Behavioral economics. In U. Mäki (Ed.), *Philosophy of Economics*, Handbook of the Philosophy of Science (Vol. 13, pp. ???-???). Amsterdam: Elsevier.

Beckerman, W. (1975). *Two Cheers for the Affluent Society: A Spirited Defense of Economic Growth*. New York, NY: Saint Martin's Press.

Bernanke, B. S. (2010, May 8). Commencement address: The economics of happiness. Retrieved from http://www.federalreserve.gov/newsevents/speech/bernanke20100508a.pdf

Blackorby, C., & Donaldson, D. (1990). A review article: The case against the use of the sum of compensating variations in cost-benefit analysis. *The Canadian Journal of Economics*, *23*(3), 471-494.

Blanchflower, D. G., & Oswald, A. J. (2004). Well-being over time in Britain and the USA. *Journal of Public Economics*, *88*(7-8), 1359-1386. doi:10.1016/S0047-2727(02)00168-8

Borsboom, D. (2005). *Measuring the Mind: Conceptual Issues in Contemporary Psychometrics*. Cambridge: Cambridge University Press.

Bridgman, P. W. (1927). *The Logic of Modern Physics*. New York: The Macmillan Company.

Camerer, C. F., & Loewenstein, G. (2004). Behavioral economics: Past, present, future. In C. F. Camerer, G. Loewenstein, & M. Rabin (Eds.), *Advances in Behavioral Economics* (pp. 3-51). New York, NY and Princeton, NJ: Russell Sage Foundation and Princeton University Press.

Campbell, A. (1976). Subjective measures of well-being. *American Psychologist*, *31*(2), 117-124. doi:10.1037/0003-066X.31.2.117

Cantril, H. (1965). *The Pattern of Human Concerns*. New Brunswick, NJ: Rutgers University Press.

Clark, A. E., & Oswald, A. J. (2002). A simple statistical method for measuring how life events affect happiness. *International Journal of Epidemiology*, *31*(6), 1139-1144. doi:10.1093/ije/31.6.1139

Cliff, N. (1992). Abstract measurement theory and the revolution that never happened. *Psychological Science*, *3*(3), 186-190. doi:10.1111/j.1467-9280.1992.tb00024.x

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281-302. doi:10.1037/h0040957

Dawes, R., & Smith, T. L. (1985). Attitude and opinion measurement. In G. Lindzey & E. Aronson (Eds.), *Handbook of Social Psychology* (Vol. 1, pp. 509-566). New York, NY: Random House.

Diener, E. (2000). Subjective well-being. The science of happiness and a proposal for a national index. *American Psychologist*, *55*(1), 34-43. doi:10.1037/0003-066X.55.1.34

Diener, E. (2006). Guidelines for national indicators of subjective well-being and ill-being. *Applied Research in Quality of Life*, *1*(2), 151-157. doi:10.1007/s11482-006-9007-x

Diener, E., Lucas, R., Schimmack, U., & Helliwell, J. (2009). *Well-Being for Public Policy*. New York, NY: Oxford University Press.

Diener, E., & Seligman, M. E. P. (2004). Beyond money: Toward an economy of well-being. *Psychological Science in the Public Interest*, *5*(1), 1-31. doi:10.1111/j.0963-7214.2004.00501001.x

Diener, E., & Suh, E. M. (1997). Measuring quality of life: Economic, social, and subjective indicators. *Social Indicators Research*, *40*(1-2), 189-216. doi:10.1023/A:1006859511756

Diener, E., Suh, E. M., Lucas, R. E., & Smith, H. L. (1999). Subjective well-being: Three decades of progress. *Psychological Bulletin*, *125*(2), 276-302. doi:10.1037/0033-

2909.125.2.276

Dupuit, J. (1969). On the measurement of public works. In K. J. Arrow & T. Scitovsky (Eds.), *Readings in Welfare Economics* (pp. 255-283). Homewood, IL: R. D. Irwin.

Durlauf, S. N., & Blume, L. (Eds.). (2008). *The New Palgrave Dictionary of Economics* (2nd ed.). New York, NY: Palgrave Macmillan.

Eatwell, J., Milgate, M., & Newman, P. (Eds.). (1998). *The New Palgrave: A Dictionary of Economics* (Corrected paperback ed.). London: Macmillan.

Frey, B. S. (2008). *Happiness: A Revolution in Economics*. Cambridge, MA: MIT Press.

Frey, B. S., & Stutzer, A. (2002). *Happiness and Economics: How the Economy and Institutions Affect Well-Being*. Princeton, NJ: Princeton University Press.

Green, J., & Hojman, D. (2007, November). *Choice, rationality and welfare measurement*. Harvard Institute of Economic Research Discussion Paper No. 2144; KSG Working Paper No. RWP07-054, . Retrieved from http://ssrn.com/abstract=1030342

Griffin, J. (1986). *Well-Being: Its Meaning, Measurement, and Moral Importance*. Oxford: Clarendon Press.

Gurin, G., Veroff, J., & Feld, S. (1960). *Americans View Their Mental Health: A Nationwide Interview Survey*. New York, NY: Basic Books.

Hart, H. (1940). *Chart for Happiness*. New York, NY: Macmillan.

Hicks, J. R. (1943). The Four Consumer's Surpluses. *The Review of Economic Studies*, *11*(1), 31-41.

John, O. P., & Benet-Martínez, V. (2000). Measurement: Reliability, construct validation, and scale construction. In H. T. Reis & C. M. Judd (Eds.), *Handbook of Research Methods in Social and Personality Psychology* (pp. 339-369). New York, NY: Cambridge University

Press.

Johnson, J. A. (2001). Personality psychology: methods. In N. J. Smelser & P. B. Baltes (Eds.), *International Encyclopedia of the Social & Behavioral Sciences* (Vol. 16, pp. 11313-11317). Oxford: Pergamon.

Judd, C. M., & McClelland, G. H. (1998). Measurement. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The Handbook of Social Psychology* (4th ed., Vol. 1, pp. 180-232). New York, NY: Oxford University Press.

Just, R. E., Hueth, D. L., & Schmitz, A. L. (2004). *The Welfare Economics of Public Policy: A Practical Approach to Project and Policy Evaluation*. Cheltenham: Edward Elgar.

Kahneman, D. (1999). Objective happiness. In D. Kahneman, E. Diener, & N. Schwarz (Eds.), *Well-Being: The Foundations of Hedonic Psychology* (pp. 3-25). New York, NY: Russell Sage Foundation.

Kahneman, D. (2003). A psychological perspective on economics. *The American Economic Review*, *93*(2), 162-168. doi:10.1257/000282803321946985

Kahneman, D., & Krueger, A. B. (2006). Developments in the measurement of subjective well-being. *Journal of Economic Perspectives*, *20*(1), 3-24. doi:10.1257/089533006776526030

Kahneman, D., Krueger, A. B., Schkade, D., Schwarz, N., & Stone, A. (2004a). Toward national well-Being accounts. *The American Economic Review*, *94*(2), 429-434. doi:10.1257/0002828041301713

Kahneman, D., Krueger, A. B., Schkade, D. A., Schwarz, N., & Stone, A. A. (2004b). A survey method for characterizing daily life experience: The Day Reconstruction Method. *Science*, New Series, *306*(5702), 1776-1780. doi:10.1126/science.1103572

Kahneman, D., Wakker, P. P., & Sarin, R. (1997). Back to Bentham? Explorations of

experienced utility. *The Quarterly Journal of Economics*, *112*(2), 375-405. doi:10.1162/003355397555235

Kline, P. (1998). *The New Psychometrics: Science, Psychology, and Measurement*. London: Routledge.

Korsgaard, C. M. (1993). G. A. Cohen: Equality of what? On welfare, goods and capabilities. Amartya Sen: Capability and well-being. In M. C. Nussbaum & A. Sen (Eds.), *The Quality of Life* (pp. 54-61). Oxford: Clarendon Press.

Krantz, D. H. (1972). Measurement structures and psychological laws. *Science*, *175*(4029), 1427-1435. doi:10.1126/science.175.4029.1427

Krantz, D. H. (1991). From indices to mappings: The representational approach to measurement. In D. R. Brown & J. E. K. Smith (Eds.), *Frontiers of Mathematical Psychology: Essays in Honor of Clyde Coombs* (pp. 1-52). New York, NY: Springer.

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Additive and Polynomial Representations*. Foundations of Measurement (Vol. 1). New York, NY: Academic Press.

Krueger, A. B. (Ed.). (2009). *Measuring the subjective well-being of nations: National accounts of time use and well-being*. Chicago, IL: The University of Chicago Press.

Layard, P. R. G. (2005). *Happiness: Lessons from a New Science*. New York, NY: Penguin Press.

Lyubomirsky, S., & Lepper, H. S. (1999). A measure of subjective happiness: Preliminary reliability and construct validation. *Social Indicators Research*, *46*(2), 137-155. doi:10.1023/A:1006824100041

Mäki, U. (1998). Realisticness. In J. B. Davis, D. W. Hands, & U. Mäki (Eds.), *The Handbook of Economic Methodology* (pp. ???-???). Northampton, MA: Edward Elgar.

Mandler, M. (1999). *Dilemmas in Economic Theory: Persisting Foundational Problems of Microeconomics*. New York, NY: Oxford University Press.

Mankiw, N. G. (2001). *Principles of Microeconomics* (2nd ed.). Fort Worth, TX: Harcourt College.

Marshall, A. (1920). *Principles of Economics: An Introductory Volume* (8th ed.). London: Macmillan.

Mas-Colell, A., Whinston, M. D., & Green, J. R. (1995). *Microeconomic Theory*. New York, NY: Oxford University Press.

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

Nordhaus, W. (2009). Measuring real income with leisure and household production. In A. B. Krueger (Ed.), *Measuring the subjective well-being of nations: National accounts of time use and well-being* (pp. 125-144). Chicago, IL: University of Chicago Press.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory* (3rd ed.). New York, NY: McGraw-Hill.

Nussbaum, M. C., & Sen, A. (Eds.). (1993). *The Quality of Life*. Oxford: Clarendon Press.

Okasha, S. (2002). *Philosophy of Science: A Very Short Introduction*. Very short introductions ;; 67; Variation: Very short introductions ;; 67. Oxford: Oxford University Press.

Pigou, A. C. (1952). *The Economics of Welfare* (4th ed.). London: Macmillan.

Rabin, M. (2002). A perspective on psychology and economics. *European Economic Review*, *46*(4-5), 657-685. doi:10.1016/S0014-2921(01)00207-0

Rubinstein, A., & Salant, Y. (2008). Some thoughts on the principle of revealed preference. In A. Caplin & A. Schotter (Eds.), *The Foundations of Positive and Normative Economics: A*

Handbook (pp. 115-124). New York, NY: Oxford University Press.

Salmon, W. C. (1992). Scientific Explanation. In M. H. Salmon, J. Earman, C. Glymour, J. G. Lennox, P. Machamer, J. E. McGuire, J. D. Norton, et al. (Eds.), *Introduction to the Philosophy of Science* (pp. 7-41). Englewood Cliffs, NJ: Prentice Hall.

Scanlon, T. (2000). *What We Owe to Each Other*. Cambridge, MA: Belknap Press.

Scott, D., & Suppes, P. (1958). Foundational aspects of theories of measurement. *The Journal of Symbolic Logic*, *23*(2), 113-128.

Sen, A. (1982). *Choice, Welfare, and Measurement*. Cambridge, MA: MIT Press.

Slesnick, D. T. (1998). Empirical approaches to the measurement of welfare. *Journal of Economic Literature*, *36*(4), 2108-2165.

Tiberius, V. (2006). Well-being: Psychological research for philosophers. *Philosophy Compass*, *1*(5), 493-505. doi:10.1111/j.1747-9991.2006.00038.x

Trout, J. D. (1998). *Measuring the Intentional World: Realism, Naturalism, and Quantitative Methods in the Behavioral Sciences*. New York, NY: Oxford University Press.

Watson, G. (1930). Happiness among adult students of education. *Journal of Educational Psychology*, *21*(2), 79-109. doi:10.1037/h0070539

Wilson, W. R. (1967). Correlates of avowed happiness. *Psychological Bulletin*, *67*(4), 294-306. doi:10.1037/h0024431

Winter, D. G., & Barenbaum, N. B. (1999). History of modern personality psychology theory and research. In L. A. Pervin & O. P. John (Eds.), *Handbook of Personality: Theory and Research* (2nd ed., pp. 3-27). New York, NY: Guilford Press.