CrossMark

RESEARCH PAPER

Happiness Studies and the Problem of Interpersonal Comparisons of Satisfaction: Two Histories, Three Approaches

Shiri Cohen Kaminitz[1]

Abstract An old methodological obstacle confronts the use of Life Satisfaction surveys in Happiness Studies: a problem that economists recognize by the name of (the impossibility of) interpersonal comparisons of satisfaction/utility. But the recent implementation of insights from happiness studies into policy making transforms an originally theoretical obstacle into a real-world problematic, providing substantial motivation for engaging with this issue. Just this problem is highlighted by recent critics of happiness surveys. This paper locates the problem currently facing happiness studies at the intersection of two traditions or two histories: that of economic methodology and that of psychological methodology. Three dominant approaches to the issue revealed through these histories are identified: 'the skeptical approach', 'the pragmatic approach', and 'the ethical-normative approach'. The paper works to bring together the two disciplinary histories and evaluate the three approaches in order to frame a suitable interpretation of inter-personal comparisons in happiness studies. The implications of this are twofold: it contributes to the legitimation of happiness studies, suggesting an answer to its critics, while, at the same time casting the status of its building blocks under a different light.

Keywords Interpersonal comparisons · Life satisfaction · Methodology · Happiness · Utility · Psychology · Economics

1 Introduction

In the past few decades, happiness studies, or the science of subjective well-being (SWB), has become a flourishing field and a shared playground for psychologists, economists and other theorists and social scientists. However, an old methodological obstacle confronts it: a problem that economists recognize by the name of (the impossibility of) interpersonal

✉ Shiri Cohen Kaminitz
  Shiri.cohen2@mail.huji.ac.il

[1] Hebrew University, Jerusalem, Israel

comparisons of satisfaction/utility (IPCS).[1] This fundamental methodological question has not as yet received a sufficient treatment in the context of happiness studies. In particular, it has not yet been treated in a way that takes into consideration the distinctive inter-disciplinary context in which this time-honored problem is now re-emerging. In addition, today, more than ever before, scholars are attempting to integrate insights and data from happiness studies into policy making,[2] and garnering the interest of both politicians and civil servants in so doing. The implementation of insights from happiness studies into policy making transforms an originally theoretical obstacle into a real-world problematic, and provides a substantial motivation for engaging with this issue.

Under the umbrella of 'happiness studies' and the science of SWB are found a number of conceptions and methodological stances toward the collection and analysis of data. The particular practice of concern to this paper, however, is that of Life Satisfaction (LS) surveys. These are based on the collection of data: large-scale surveys of self-reports in which people ascribe a score to their level of satisfaction from their 'life as a whole' within a given scale.[3] This data is then analyzed with the aid of statistical and econometrical tools, controlling for suggested factors and searching for interesting correlations with other factors. As stated, this is not the only method in happiness studies, but it is a common one, and widely used, so that it can provide a representative case. Moreover, this particular methodology in happiness studies is the most vulnerable to the criticism of IPCS. Thus, an illustration of the basic problem addressed here might be as follows: in the case of a 1 to 7 scale of life-satisfaction degrees, just what, if anything, guarantees that the grade of "6" of different individuals denotes a similar content, or that person A's "4" expresses something more than B's "3"? While some proponents of happiness studies address this issue explicitly yet find nothing within it especially problematic (e.g. Veenhoven (2010) and Ng (1996)), just this problem is highlighted as particularly damning by recent critics of happiness surveys such as Adler (2012) and the more caustic McCloskey (2012).

This paper addresses this issue by way of a recognition of the interdisciplinary nature of contemporary happiness studies. In particular, the problem is located at the intersection of two traditions, or two histories: that of economic methodology and that of psychological methodology. A dialogue between these two disciplines promises to be fruitful, for each has significant relative advantages: psychologists have much broader experience with the methodologies of self-scaling and emotions-measurements, while economists have experience in explicitly tackling the problem of IPCS. By identifying previous ways of engaging with the problem, noting their positive features as well as their pitfalls, it becomes possible to establish a suitable framing of—and hence offer a suitable approach to resolving—the current, unique version of the problem.

It should be stressed that the various sections of the paper by no means offer a comprehensive survey of the history of the problem of IPCS in economics[4] or related problems

---

[1] I choose to use the term satisfaction and not utility here, first, in order to avoid confusions about the changing meaning of the concept of utility in the history of economics, and, second, because it is the term more frequently met in the language of psychologists.

[2] Prominent examples include Oswald (1997), Diener and Seligman Martin (2004), Layard (2005), Layard (2006), Diener (2009), Veenhoven (2010), Frey and Stutzer (2009), Frey and Stutzer (2012).

[3] For example, researchers widely use surveys such as World Values Surveys in which the one item (question) is: "All things considered, how satisfied are you with your life as a whole these days? Please use this card to help with your answer [range of 1–10 with 1 labelled "Very Dissatisfied" and 10 labelled "Very Satisfied"]. For more examples see Adler (2012, pp. 2–3).

[4] Extensive surveys of the subject in the history of economics has been provided by Sen (1979), Hammond (1991), Fleurbaey and Hammond (2004).

in psychology. Rather, a modest attempt is made to provide a general categorization of prominent traditional approaches to the problem, and to use this categorization as a guide to the most suitable approach in the current context.

An attempt to facilitate dialogue between economists and psychologists entails some terminological and even conceptual redrafting. A basic distinction has been made over the years by economists between descriptive and normative interpretations of IPCS.[5] This paper employs a related but different distinction that may better clarify the issue at hand, identifying three dominant approaches to the problem, which are here labelled the skeptical approach, the pragmatic approach, and the ethical-normative approach.

The first section of the paper presents the skeptical approach, held by many economists deep into the second half of the twentieth century. These economists presented the problem of IPCS in a persuasive and influential manner, characterizing it, in the spirit of logical positivism, as unscientific. Section 1.2 offers some necessary analytical clarifications concerning the special case of IPCS in happiness studies, which follow from our presentation of the skeptical approach.

The second section addresses the pragmatic approach, which was developed in the second half of the twentieth century and which is the dominant approach adopted today within happiness studies and, in particular, in the practice of interpreting LS surveys. The pragmatic approach encompasses a wide variety of methods, implemented by both economists and psychologists in order to legitimize IPCS. All such approaches are characterized, according to the terminology used in this paper, by the ascription of a scientific status to IPCS (or to related concepts) and the avoidance of any explicit ethnical-normative grounds. Two particular pragmatic approaches are singled out for inspection, both with roots in the 1950s, but the one developed by economists (Sect. 2.1) and the other by psychologists (Sect. 2.2). These two approaches are particularly relevant for addressing the current state of play of IPCS in life-satisfaction data.

Finally, the third section identifies the pitfalls of the pragmatic approach in the context of implementing happiness studies in public policy and looks to the ethical approach as providing a remedy to these shortcomings. The case made for the ethnical approach builds upon the analysis of Fleurbaey and Hammond (2004) and argues for the need to acknowledge that IPCS made within happiness studies are based on ethical-normative presumptions.[6]

The paper as a whole aims to contribute to the legitimation of the implementation of happiness studies within public policy, suggesting an answer to its critics on this specific point, while, at the same time casting the status of the building blocks of happiness studies in a different light. The conclusions of the paper are significant for the way social scientists understand their project, and important, in particular, for the comprehension of happiness studies by politicians and by the general public.

1.1 The Skeptical Approach

The skeptical approach to IPCS has deep roots in the history of economics, and has been crucial ever since utility functions were introduced into economic theory. As soon as

---

[5] See Fleurbaey and Hammond (2004, pp. 46–49), Davis (1992), Rosenbaum (1995).

[6] I relate here to Hands's important clarification of the distinction between the mere 'normative' versus the ethical-normative (Hands 2012). The later constitutes a case of dependence on values and ethical convictions, i.e. not just on any kind of value judgement, but on one that relates to social arrangements, welfare, fairness or justice.

microeconomics was established upon individual utility functions the question became inevitable as to the value of such mathematical formalizations when social utility was in question. Thus, already in the 1870s, the first generation marginalist, William Stanley Jevons,[7] insisted that:

The reader will find… that there is never, in any single instance, an attempt made to compare the amount of feeling in one mind with that in another. I see no means by which such comparison can be accomplished… Every mind is thus inscrutable to every other mind, and no common denominator of feeling seems to be possible.[8]

The problem was highlighted in passing as twentieth-century conceptual schemes and methodological approaches evolved. In particular, the problem was raised acutely when utility was considered as an introspective entity, which happened when utility was considered a function representing revealed preferences only. This occurred when the differential calculus became the main tool of economics, and again under the axiomatic/representational methodological approach; both under cardinalist and ordinalist interpretations of the utility functions.[9]

The skeptical approach to the problem of IPCS was explicitly posed by economists in the first half of the twentieth century, most famously in Lionel Robbins's *Essay on the Nature and Significance of Economic Science* (1932). Robbins's highly influential conceptualization of the problem can usefully be approached in terms of the dominant philosophical position of the day, namely logical positivism. In accordance with the spirit of the time, in economics, as in other sciences, a demarcation line was demanded between the meaningful and meaningless, the scientific and the 'unscientific' (the metaphysical, the normative, etc.). IPCS fell outside the demarcated line and range of scientific economics— a view generally accepted within the economics community at the time and for decades afterwards. The main argument presented in Robbins's celebrated essay was that the interpersonal comparison of satisfactions is *unscientific*:

It is a comparison which necessarily falls outside the scope of any positive science. To state that A's preference stands above B's in order of importance is entirely different from stating that A prefers n to m and B prefers n and m in a different order. It involves an element of conventional valuation. Hence it is essentially normative. It has no place in pure science.[10]

---

[7] Enthused by the doctrine of utilitarianism and by its psychological-hedonist presumptions aimed at combining utilitarian elements with cutting edge economic theory. The result in Jevons's case was the formulation of a mathematical representation of individual utility (focusing on the 'lower feelings'). Jevons, nevertheless, avoided using this as a platform for a mathematical formulation of social aggregate utility/satisfaction.

[8] Jevons (1888), paragraph 20. Jevons's colleague, Edgeworth, by contrast, did not hesitate to do exactly this. Under what he titled 'exact utilitarianism,' Edgeworth presented abstract problems of allocation (of stimulus, and then of 'means to stimulus,' such as wealth and labor) using basic functions of pleasure-producing, so that the aggregate of pleasures/happiness, the result of that allocation, would be at a maximum.

Edgeworth (1877, 1879). The basic mathematical representation was:

$$\int \int \int (dp)(dt)(dn)$$

where p is pleasure degree, t is time duration, n is 'number of enjoyers.' See Edgeworth (1879, p. 394).

[9] The interpretation of utility as cardinal ascribes to the utility function more than just an ordering of preferences of one individual. (i.e. it addresses utility as a magnitude).

[10] Robbins (1934, p. 123).

Interestingly, psychologists' reaction to logical positivism led them down a completely different path, as will be discussed in Sect. 2.2.

A few issues should be highlighted when considering Robbins's point of view. Firstly, and significantly, Robbins made it clear that shoving IPCS outside the accepted range of economic science was *not* necessarily the result of the influence of behaviorism in economics. If we focus on *behaviorism* versus *introspection* in economics, the latter supposes that economic utility functions can tell us something about the internal world of individuals, as opposed to the former, which denies the scientific status of such an attempt. Robbins was far from being a strict behaviorist,[11] and his objection to IPCS was based rather on his acceptance of introspection as a legitimate assumption. Thus, in a later paper:

I still cannot believe that it is helpful to speak as if interpersonal comparisons of utility rests upon scientific foundations—that is, upon observation *or introspection*.[12]

One can view the methods developed today within happiness studies as marking a shift from revealed preference method in economics back to introspection (by way of self-report surveys). In the context of this paper, then, Robbins's position recalls the basic intuitive conviction that the IPCS problem constitutes a serious obstacle for both methodological stances: behaviorism and introspection. This point is crucial in order to place the current methodological problem on the same line as the old one. Thus, the skeptical approach regarding IPCS can be relevant both when behaviorist methodologies are operated as well as when introspective methodologies are operated.

A second point about Robbins's view that is crucial for our purposes is that his rejection of IPCS is based on his conception of the demarcation line defining the 'scientific.' What Robbins actually means is that IPCS presumes an ethical-normative view,[13] and one cannot base a science on contingent ethical convictions. Ironically, the basis of this objection also opens the possibility of overcoming the problem in two possible ways: one is to argue that IPCS is a non-normative judgment; the other is to declare the demarcation line unnecessary and to allow the ethical-normative and the scientific to become entangled (i.e. to let go of positivism). These two possibilities will be manifested in our historical discussion of the two other approaches to IPCS: the 'pragmatic approach' and the 'ethical approach.'

1.2 Posing the Problem Within Happiness Studies

Before proceeding, and in order to further focus on the nature of the current problem of IPCS and life-satisfaction, three analytical clarifications are required.

Firstly, mid-twentieth economists were dealing with 'utility' and not with 'life-satisfaction.' In important respects the analogy between the two is valid, but the differences should be noticed and are helpful in clarifying the nature of the problem. One difference is the methodological gap, or the question of *how* we receive and handle information: real-introspection or self-report was obviously not then a common methodology within economics for forming (hypothetical-deductive) utility functions. In addition, the scientific tools and practical methodologies used in both cases are completely different.[14]

---

[11] See Hands (2009).

[12] Robbins (1997).

[13] See footnote 6 above.

[14] See Angner (2009a) on the two methodologies: the psychometric approach used in happiness studies, and the axiomatic/representative approach in welfare economics. For the basic distinction see Krantz (1991).

A related question concerns what we are trying to capture and compare. What is utility? Fumagalli recently summarized the three interpretations of utility made by economists.[15] That most established regards utility as a *mathematical representation of preferences* (*decision utility*).[16] The second refers to some purported hedonic magnitude reflecting individuals' experiences of pleasure and pain (*experienced utility*).[17] The third takes utility as a desirability signal that can be accurately measured in the activation patterns of specific neural areas (*neural utility*).[18] What do life- satisfaction surveys capture? Or what kind of 'utility' do they resemble? Obviously, and unfortunately, they do not resemble the third interpretation (neural utility).[19] As emphasized by Adler (2012), it does not resemble *experienced utility* either, because people's reflection on their life-satisfaction is *not* purely 'experiential' (it is not a reflection of mental states alone).[20] In some respects it resembles *decision utility* (or as Adler calls it: preference-based account), albeit with the methodological gap between the two.

In this paper, therefore, the question concerning IPCS in LS data cannot be approached through solutions that takes happiness to be a mere mental state (as in Kahneman's experienced utility),[21] or a mere neural description. Life-satisfaction surveys describe a different interpretation to happiness than these alternatives, namely a *deliberative* (and not only emotional) introspective assessment of one's satisfaction with one's life as a whole.

Another analytical clarification turns on the fact that in focusing on collecting and analyzing life-satisfaction data there are two basic aspects suspected of being ethically normative. Firstly, the aspect of using the data (the many numbers representing levels of satisfaction) in a particular way: computing averages, adding up the numbers etc. In the case of happiness studies the particular way of approaching the numerical values is similar to a utilitarian social function, i.e. ascribing the individual values with equal weights.[22] This is obviously a methodological aspect that is ethical-normative. But this is not exactly what is at issue. It is the former aspect that might make the difference between the different approaches to the question of IPCS: the very meaning of the data itself includes an assumption about IPCS (the *first* aspect). What is the status of this information? Even

---

[15] Fumagalli (2013, pp. 325–326).

[16] Broome (1991), Hausman and McPherson (2009).

[17] Kahneman et al. (1997), Kahneman et al. (1999), Kahneman and Krueger (2006).

[18] Camerer et al. (2004).

[19] It is unfortunate because neural utility is, perhaps, the most resistant to the IPCS problem.

[20] Adler (2012, pp. 10–14). The interpretation of LS data as non-purely-experiential goes hand in hand with psychologists' distinction between affective *and* cognitive components of subjective well-being and their designing scales in order to reflect both. Life-satisfaction scales were implicitly designed as more than mere affective-well-being scales. See Lyubomirsky and Lepper (1999), Diener (1994).

[21] Kahneman and Krueger, who have taken Robbins's economic point of view explicitly into account. Thus they make use of this very problem when arguing for the need for alternating the methodology of life-satisfaction surveys with other methodologies for the measurement of subjective well-being (such as the U-index): "One of the difficulties of using data on subjective well-being is that individuals may interpret and use the response categories differently… when Tim answers a 4 about the intensity of a particular emotion, maybe that is the equivalent of a 6 for Jim… We propose an index, called the U-index which overcomes this problem." Kahneman and Krueger (2006, pp. 18–19). Whether or not these scientists succeeded in overcoming with their suggested methodology the problem of IPCS is an interesting question, albeit outside the scope of this paper.

[22] For this point see also Angner (2009b, pp. 158–163), Fleurbaey and Hammond (2004, p. 52).

before adding up (no matter how), the question arises on what basis we make the assumption that the comparable numbers represent comparable levels of satisfaction?[23]

Finally, focusing on the first aspect and trying to locate the IPCS element within it: addressing the data itself as meaningful (before even analyzing it, etc.) implicitly entails the making of two assumptions: (1) *on the intra-personal level*, that individuals actually have an 'access' or awareness of their levels of life-satisfactions, and that the differences between the scores are the same for each individual (although not necessarily between individuals)[24]; (2) the normalizing of the scales, that is, assuming that the different scales of different individuals share the same bottom and upper levels. These two assumptions may pave the way to IPCS.

In case one accepts the view that IPCS is indeed a problem that should be addressed in this case, then, the basic question becomes on what basis this hypothetical calibration is to be made? And more particularly, does it involve a normative-ethical judgement?

2 Pragmatic approaches

2.1 Harsanyi's Pragmatic Approach

Economics as a twentieth-century discipline, heavily influenced as it was by the skeptical approach, could in most cases manage without IPCS so far as micro and macro theory were concerned; and the very possibility of doing well without it seems to have strengthened the strategy of dispensing with it altogether. Economics between the 1930s and the 1950s thus managed to avoid using IPCS.[25] This was at first so even in the case of welfare economics (called by then "new welfare economics"). Nevertheless, within this sub-discipline, unless one restricts social welfare functions only to 'Pareto improvements' cases, in which all individuals are better off in a particular situation compared to another, one is compelled to make IPCS.

Not surprisingly, during the second half of the twentieth century a variety of pragmatic approaches were developed by economists to tackle the problem within welfare economics and related fields.[26] Being acutely aware of the skeptical view, many of the suggestions allowed IPCS by basing them on *descriptive* and allegedly non-normative conduct (Hammond 1991, 211–226). Some of the procedures involved cardinalisation and

---

[23] Here the approaches may differ. One could argue for the objective or descriptive status of the data. This might be done by suggesting that once the many individuals make their reports in a contemplative manner (with no hidden intra-personal problematics), there is no need for an external authority to ascribe the given scores with meaning (thereby making them useful) This is the approach endorsed in Ng (1996). A counterargument might be that this step necessarily requires an additional external ethical-normative judgment.

[24] See for example the constructive suggestions regarding these issues in Fleurbaey and Blanchet (2013, pp. 199–201). Their suggestion focuses on the particular design of the questionnaires. Another attempt to face these challenges is Ng (1996). Whether this problem is solvable is an interesting question, but beyond the scope of this paper since this is not the stage involving IPCS.

[25] Hence the development of a microeconomics that connected competitive market solutions with Pareto optimality (via the first and the second 'fundamental theorems') and that was therefore, allegedly, immune to IPCS.

[26] See the surveys by Sen (1979) and Hammond's (1991). Among those approaches a very partial list would include: Harsanyi (1955), Little (1950), Waldner (1972), Ng (1975, 1982). See also Davis (1992) for a philosophical account of the issue. Davis characterized inter personal comparisons of utility as descriptive and *value-laden* and so not *ethical*-normative.

normalizing of scales of utilities[27] in a way that might seem analogous to the conduct of calibration implicit in happiness studies (see the previous section). However, in welfare economics, as in happiness studies, a question should be raised concerning the grounds on which this calibration is conducted.[28]

Only part of the solutions raised by economists during this phase are actually relevant to the concrete problem of IPCS within happiness studies. This is primarily because of the methodological gap between happiness studies and welfare economics. As explained, the problem faced in this paper is much more concrete than the general question of allowing for IPCS in all the fields of economics. Addressing one particular strategy, which seems more relevant to our particular problem, might be constructive. John C. Harsanyi's conceptualization of the issue, which took IPCS to be based on 'inductive logic,' provides just such an early influential strategy.

In his seminal essay, 'Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility' (1955), Harsanyi explicitly addressed the problem by distinguishing between two aspects of the problem (on which, see the section above). He first addressed the procedure of weighing individual utilities in a particular manner (in a utilitarian social function, in this case). This part, which takes place only *after* IPSC is conducted, is described as *ethical* by essence.[29] It should be emphasized, though, that this is a separate issue from the basic IPCS problem that is dealt with in the closing section of the essay. Within this closing section Harsanyi addresses directly the challenge of IPCS and demands an analysis of its "logical basis."

As opposed to the ethical nature of the social welfare function, the comparison between individual utilities is considered by Harsanyi as *logical* and not (necessarily) ethical. In short, Harsanyi set the ground for curbing the problem by suggesting a basic distinction between the *metaphysical* question and the practical one, which he termed the *psychological* question.[30] While the metaphysical question must remain always unresolved because the scientist can *never* know whether one's satisfaction is *really* bigger or smaller than another's, no matter how detailed the indications given (whether by self-reports or by revealed preferences), it is a different case with the psychological-practical question.[31]

The task becomes one of identifying the variables relevant to satisfaction/utility and framing the laws based on empirical investigation. Then, armed with as complete as possible knowledge of such, one can make IPCS, contingent on those laws; whereas:

In general, the greater the psychological, cultural, and social differences between two people, the greater the margin of error attached to comparisons between their utility.[32]

Utility, therefore, is considered as the unexplained remainder, after the scientist controls for other variables. Although such an empirical task might seem endless and never perfect,

---

[27] Economists such as Isbell (1959) and Schick (1971) have suggested such strategies, i.e. first cardinalization by using preferences over risk (probabilities), a method used by von Neumann and Morgenstern, and then normalizing the scales by putting upper and lower bounds on all individual utilities. See in Hammond (1991, pp. 215–216).

[28] As it raised by Hammond, ibid. p. 216.

[29] Harsanyi (1955, pp. 315–316).

[30] Harsanyi (1955, p. 317).

[31] Described as follows: "If in a given situation one individual gives more forcible signs of satisfaction or dissatisfaction than another, is this so because the former feels more intense satisfaction… or only because he is inclined to give stronger expression to his feeling?" Ibid. p. 318.

[32] Ibid. pp. 318–319.

this phrasing of the problem makes IPCS logical *in theory*. As Harsanyi puts it (in opposition to Robbins[33]):

…it should now be sufficiently clear that interpersonal comparisons of utility are not value judgments based on some ethical or political postulates but rather are factual propositions based on certain principles of inductive logic.[34]

This approach could easily be implemented in the case of IPCS in happiness studies by way of the following assumption: the more the subjects of our investigation share in common their social, cultural, psychological, educational (etc.) situations, the more solid is the logical basis upon which the inter-personal comparisons are made.[35]

Harsanyi's approach is important and the various later versions derived from it share the view that: (1) in economics, as in day-to-day life, interpersonal comparisons of utility are made as a matter of fact; (2) the task of the social scientist or the philosopher is to explain the basic intellectual operation underlying such a fact; (3) an explanation does exist, and is not necessarily one that includes ethical judgements (contrary to the method of weighing utilities in a social welfare function).

Acceptable or not in the various sub-disciplines of economics,[36] the question at issue here is whether this account is sufficient in the context of IPCS in contemporary happiness studies and its implementation in public policy. In Sect. 3 it will be argued that it is not, and that an extra postulate is needed—this being an ethical-normative postulate. Before we proceed to this discussion, however, we must turn to the path walked by the psychologists.

2.2 The Psychologists' Pragmatic Approach

Daniel Kahneman and his associates commented in a recent paper that psychologists "are more comfortable than economists when it comes to comparing indicators of feelings or utility across individuals".[37] I believe this to be correct and reflects the pragmatic approach adhered to by psychologists. In this section some important landmarks in the evolution of the pragmatic approach will be presented and discussed, as also its relevancy to and consequences for the issue of IPCS. The problem of IPCS is related to central methodological concerns conceptualized (such as the concept of *measurement*, the validation of scales, etc.) that psychologists conceptualize differently to economists.

A recent and thorough overview (Angner 2011) reminds us of the long history within psychology of attempts to measure happiness and satisfaction. This history commences back in the 1920s and 1930s, with various studies of education, personality and marital success. Nevertheless, and as pointed out by Kahneman, Diener and Schwartz, the "study of hedonics" could not thrive under the intellectual regimes of logical positivism, behaviorism and the cognitive revolution, for "it could not be elegantly described in the

---

[33] It should be noted, though, that Harsanyi does agree with Robbins that *in some cases* IPCS does involve "ethical or political restrictive postulates", but holds that those cases should be distinguished from the cases of IPCS without conventional element of this kind. Ibid. p. 320.

[34] Ibid pp. 319–320.

[35] Harsanyi's later and more famous accounts of inter-personal comparisons, which used such concepts as 'the *similarity postulate*,' '*imaginative empathy*' and '*extended preferences*,' are in line with this approach. Nevertheless, they do not add much to the basic solution offered to our concerns here. Harsanyi (1982, p. 50). This is also the case with the spectator's 'extended preferences'; see Adler (2014, pp. 126–131).

[36] A recent account of the downfalls of this approach in welfare economics is given by Adler (2014). Adler suggests to base inter-personal comparisons of utility on spectator's *sympathy*. As stated by Adler this alternative base, contrary to Harsanyi's 'imaginative empathy,' is not value-free. Ibid p. 150.

[37] Kahneman et al. (2004, p. 432).

dominant theoretical language of the day, and as a consequence it was relegated to peripheral regions."[38]

Nevertheless, approaching the history of the problem from the perspective of psychologists reveals that they have been measuring a wide range of "attributes" (sensations, traits, cognitive abilities etc.) from at least the early nineteenth century.[39] In particular, in past decades one method used by psychologists is that of self-report scales ('Likert scales') for a variety of attributes and subjective experiences. So the use of scaling of happiness/ satisfaction and interpreting the data as meaningful (i.e. comparing the satisfaction-scores between distinct individuals) is only a particular case of a more general practice of measuring and comparing between all kinds of scores given by distinct individuals. Thus, while we must here attend to the history of happiness research in psychology, the more important history is that of *measurement* and scales in general, within which happiness-scales are but a particular case.

An interesting starting point is the reaction of psychologists to logical positivism in the 1930s and 1940s, the period of intellectual history that saw economists embracing the skeptical approach to the question of scientific comparisons of utility between individuals. Significantly, the positivistic influence lead psychologists along a completely different path. In particular, the work of S.S. Stevens, which was conducted as a reaction to logical positivism, actually opened the door to quantified scientific comparisons between individuals' inner worlds.[40] The key to unlocking this door was Stevens's seminal definition of measurement (1946):

> measurement, in the broadest sense, is defined as the assignment of numbers to objects or events according to rules.[41]

This is what philosophers of science would call a *nominalist* definition of measurement. For Stevens, methods of measurement are definitive of concepts; a view that stands in opposition to *realism*, which that takes measurements to be methods of finding out about objective quantities that we can identify independently of measurement.[42] Indeed, Stevens's nominalism took the radical form of *operationalism*[43]: the view that the meaning of a concept is fully specified by its method of measurement, implying that each measurement operation *defines* its own concept.[44] The relevancy of this stance to our concerns is straightforward: scientists who hold to nominalism do not commit to measuring *real* entities – *real* life-satisfaction/happiness included; so from this perspective the impossibility of inter-personal comparisons of *real quantities* of satisfaction is obvious, but irrelevant ("the metaphysical question," to use Harsanyi's terminology, is to be set aside).

Stevens's definition was widely accepted within the psychological community and integrated into the basic psychological toolbox and textbooks of the second half of the

---

[38] Kahneman et al. (1999), preface p. ix.

[39] See Michell (1999, chapter 4).

[40] Stevens's idea was developed especially in reaction to skeptical views concerning psychology and scientific measurement expressed by *The Ferguson Committee* (1940); see Michell (1997, pp. 368–369; 1999, pp. 143–155).

[41] Stevens (1946, p. 677).

[42] Chang and Cartwright (2013, p. 367).

[43] On Stevens's operationalism (a view that he explicitly adhered to) see Michell (1999, pp. 169–177).

[44] The other, more moderate nominalism is known as conventionalism, according to which we are free to choose by agreement the correct measurement method for a concept; see Chang and Cartwright (2013, p. 368).

twentieth century. It has opened the way for tremendous progress in the development of methods of measurement by putting to one side substantial philosophical questions pertaining to issues of *ontology* and *epistemology*.[45]

It is not suggested here that psychologists have in general accepted nominalism; nevertheless, it is surely significant that the basic definition of measurement in psychology is nominalist. What is suggested here is that in order to fully address the question of IPCS within psychology we cannot remain satisfied with surveying *methods* of measurement, but must also search for the (at times implicit) epistemological standpoint that they carry with them.[46]

Psychometrics (the measuring of psychological attitudes, traits and abilities) has advanced significantly since the second half of the twentieth century (see Jones and Thissen (2007) for a general overview). During this period (what came to be called) *construct validity* processes in psychological tests became increasingly theoretically sophisticated. It is through the process of construct validation that all kinds of self-report scaling tests were ascribed with meaning and scientific validity. Related problems to that of IPCS were usually dealt with through this conceptual framework.

In 1955, the same year Harsanyi published his celebrated paper in an economics journal, a paper entitled 'Construct validity in psychological tests' was published in the *Psychological Bulletin*. In this paper the authors, Cronbach and Meehl, suggested types of validation procedures for a test, with their declared larger scientific mission being to establish a "construct". As Cronbach and Meehl defined it, a construct is "some postulated attribute of people, assumed to be reflected in test performance".[47] A good example of a construct is *intelligence*, and there are many alternative tests that stand as candidates for the best way to measure intelligence. Another good example of a construct is of course *happiness*. Test validation criteria aim to distinguish a bad test from a good one. This is a far from trivial task when the construct itself cannot be directly observed. As explained in a later account of validity in psychometric theory:

> To the extent that a variable is abstract and latent rather than concrete and observable (such as the rating itself), it is called a 'construct.' Such a variable is literally something that scientists 'construct' (put together from their own imagination) and which does not exist as an observable dimension of behavior.[48]

With the development of construct validity for *test theory* arose a more sophisticated method than the traditional theory (Thurstone, 1931), at the core of which stood the *reliability*[49] and *validity*.[50] This important tradition in psychometrics is referred to as

---

[45] Michell (1997, pp. 360–361). The ontology of a theory consists in the objects the theory assumes there to be. Epistemology is the study of knowledge and justification; see Audi (2015).

[46] Chang and Cartwright distinguish between "precision" and "accuracy," which are often confused: "*Accuracy* is a realist notion about whether measurement results agree with the true values; *precision* is a concept that is meaningful to the realist and nominalist alike, as it indicates merely how specific a measurement result is." Ibid. p. 370. So focusing only on methods could lead us to far-reaching conclusions as far as improvements in *precision* are involved but not necessarily to improvements in *accuracy*.

[47] Cronbach and Meehl (1955, p. 283).

[48] Nunnally and Bernstein (1994, chapter 3: 'Validity,' p. 85).

[49] The internal consistency and stability with which a measuring instrument performs its function. Colman (2008).

[50] The extent to which a test measures what it purports to measure, or the extent to which specified inferences from the test's scores are justified or meaningful. Colman (2008).

classical test theory or *true score theory.*[51] The validation process requires empirical investigation, and many forms of evidence are admitted. One prominent form of evidence might be discovered correlations among various measures of the same construct. It should be remembered, however, that:

> Measurement or test score validation is an ongoing process wherein one provides evidence to support the appropriateness, meaningfulness and usefulness of the specific inferences made from scores about individuals from a given sample and in a given context. As such it is not an all-or-none decision but rather a matter of degree.[52]

Note how even a high degree of validity does not indicate perfect *accuracy* of the scores (see footnote 46). Since the 1950, as B.D. Zumbo points out:

> [Psychologists] have moved from a correlation (or a factor analysis to establish "factorial validity"[53]) as sufficient evidence for validity to an integrative approach to the process of validation involving the complex weighing of various bodies, sources and bits of evidence – hence, by nature bringing the validation process squarely into the domain of disciplined inquiry and science[54]

Over the past decades, scientists within happiness studies have been engaged in empirical research within the conceptual framework of construct validity, and have begun a search for the most appropriate questionnaires to 'capture' the construct of happiness/life-satisfaction (which could therefore, to paraphrase Kahneman—see footnote 38—"thrive under the conception of measurement adopted by psychologists"). Thus, for example, in 1985 scholars presented a multi-item scale to measure life-satisfaction as a cognitive-judgmental process: In a process of careful selection, reports were designed to include five statements ('items') given to 176 participants, who graded the items scored from 1 to 7. Within the validation process, the results were compared with no less than 13 alternative previous tests and correlations systematically analyzed. In addition, results have been compared with rates given by professional interviewers.[55] Another example is the 'subjective happiness scale,' a measure suggested in 1997 that included four items and aimed to combine both *affective* and *cognitive* components.[56]

The implicit approach to IPCS addressed here is "pragmatic," since although there is no commitment to representing the same *real* magnitudes of life satisfaction of any particular

---

[51] Test theory is the archetype of a problem unique to psychological research that requires a statistical solution. Considering a test score from a statistical point of view, it is highly desirable to derive an ancillary statement of its precision. In the most basic approach to true score theory, the test score X is considered the sum of a true score T and a random error E., where $X = T+E$. The standard deviation of the errors E is a statement of the (lack of) precision, or standard error, of the test score. Jones and Thissen (2007, pp. 10–11)

[52] Zumbo (2007, 48, 56).

[53] The idea behind this form of validity was introduced by the English statistician and psychologist Charles Spearman in 1904, where he interpreted intelligence as the factor g that underlies all test items and subtests with good content validity (that is, items and subtests that appear to require intelligence) and argued that the most valid tests are those with highest loadings on the factor g. Colman (2008).

[54] Zumbo (2007, p. 72).

[55] Diener et al. (1985).

[56] "Results indicated that the Subjective Happiness Scale has high internal consistency, which was found to be stable across samples. Test–retest and self-peer correlations suggested good to excellent reliability, and construct validation studies confirmed the use of this scale to measure the construct of subjective happiness." Lyubomirsky and Lepper (1999).

pair of scorings (of two individuals), nevertheless, there is sufficient validity performance of the tests to allow for significant progress in proceeding with the scientific work.

Related themes in psychometrics, such as test designing, test equating,[57] scaling and linking,[58] have received extensive attention in recent decades, and have all been addressed in relations to issues surrounding IPCS. It is not within the scope of this paper to elaborate on these many different statistical and psychometric methods [for a comprehensive overview see Kolen and Brennan (2004)]. Nevertheless, the methods of Item Response Theory (IRT) and Rasch model must be mentioned, not only because they embody alternative statistical stances (vs. classical test theory), but also because of their seemingly different epistemological points of view (i.e. the stronger adherence to a realist conception of measurement emphasized by their advocates).

The foundations of IRT were developed in the late 1960s, and since then research in IRT has developed rapidly.[59] IRT is part of the ('latent variables') tradition of measurement providing an alternative to that of test-theory ('true scores').[60] In its basic features:

> IRT is a collection of mathematical models and statistical methods that are used to (a) analyze items and scales, (b) create and administer psychological measures, and (c) measure individuals on psychological constructs.[61]

> Three fundamentals of IRT are: *Item response functions* (IRF), *Information functions*[62] and *invariance*.[63] For items (the statements and questions rated by the individuals) on a rating scale, IRF is mathematical function describing the relation between where an individual falls on the continuum of a given construct and *the probability* that he or she will give a particular response to a scale item designed to measure that construct. In IRT a construct is called a *latent trait*… The basic goal of IRT modeling is to determine an IRF for each item on a measure. In turn IRF's are used to evaluate item quality and serve as building blocks to derive other important psychometric properties.[64]

---

[57] *Equating* is a statistical process that is used to adjust scores on different test forms so that scores can be used interchangeably; see Kolen and Brennan (2004, p. 2).

[58] *Linking* is concerned with situations in which statistical adjustments are made to scores for tests that differ in content and/or difficulty; see Kolen and Brennan (2004, p. 423).

[59] Among the first seminal works in the field are: Lord and Novick (1968), Samejima (1969), Bock and Lieberman (1970). For an historical review see Jones and Thissen (2007, pp. 12–13).

[60] Barsboom (2005, Chapters 2, 3).

[61] Reise et al. (2005, p. 95).

[62] To judge the quality of an item, one can transform the item's IRF into an item information function (a number that represents an item's ability to differentiate among people) the item provides at each trait level. Reise et al. (2005, p. 95).

[63] Invariance in IRT means two things. First, an individual's position on a latent-trait continuum can be estimated from his responses to any set of items with known IRFs, even items that come from different measures. (In contrast, in classical test-theory, item responses are aggregated to estimate a true score that is specific to that measure alone). Second, item properties, as represented by IRF, do not depend on the characteristics of a particular population (also, contrary to classical test-theory). Ibid. p. 96. This gives these kinds of model advantages in linking and equating (see footnotes 57, 58).

[64] Ibid. p. 95.

The methods of IRT have demonstrated essential advantages over classical test-theory in handling qualitative variation (the different ways that a psychological construct may be manifested by different types of people); in scaling individual differences (better metric)[65]; and in enabling a careful psychological analysis (identifying important scale properties and problems that are missed by traditional analysis).

A related methodology was developed independently in Denmark by Georg Rasch (Rasch 1960, 1961, 1966, 1977).[66] Rasch also founded the Institute of Objective Measurement, which is still active today.[67]

Recently, happiness studies scholars have also started to use IRT methods, acknowledging its advantages in evaluating and designing tests and scales. For example, O'Connor et al. (2015) presented IRT analyzation of the four-item Subjective Happiness Scale.[68] In particular, scholars have complemented other methods (such as structural equating modeling, SEM)[69] with IRT methods in order to address the problem of measurement invariance across known groups (i.e., cultures).[70] This particular problem has received much attention from happiness scholars in the past decades (see Diener and Suh, 2000) and hence been analyzed with increasingly sophisticated statistical tool and methods. The problem of measurement invariance across cultures, however, does not resolve the more basic problem of IPCS because it deals with a particular case, namely the comparison between different groups, and not with the more fundamental issue of comparison between any two individuals.

To conclude: two basic questions must be addressed. Firstly, what are the epistemological assumptions of both test-theory and the latent variables methods? To use Chang and Cartwright's terminology: is the better *precision* demonstrated by the advanced methods interpreted also as the better *accuracy* of the tests and scales (on the distinction see footnote 46)? Second, but connected to the first question, in what way are the improved conceptions and methods developed in psychometrics helpful for resolving the IPCS problem in happiness studies?

[65] "Trait-level estimates in IRT are superior to raw total scores (in classical test-theory) because: (a) they are *optimal* scaling of individual differences (i.e. no scaling can be more precise or reliable; (b) latent-trait scales have relatively better (i.e. closer to interval) scaling properties." Ibid p. 98.

[66] The model developed by Rasch and his followers is often called the "one parameter" model to contrast it with other IRT "two/multi parameters" models: its item parameter reflects *difficulty* (of the items) and does not consider any item *discrimination parameter*, this later representing how sensitive is the discrimination between respondents at all levels of the latent-trait continuum.

[67] Note the interesting definition of "objective measurement" on the institute website: "An objective measurement estimate of amount stays constant and unchanging (within the allowable error) across the persons measured, across different brands of instruments, and across instrument users. The goal of objective measurement is to produce a reference standard common currency for the exchange of quantitative value, so that all research and practice relevant to a particular variable can be conducted in uniform terms. Objective measurement research tests the extent to which a given number can be interpreted as indicating the same amount of the thing measured, across persons measured, and brands of instrument." http://www.rasch.org/.

[68] The same scale is analyzed in Lyubomirsky and Lepper (1999).

[69] SEM represents a method that combines factor analysis model (see footnote 53) with multiple regression (econometrics). Jones and Thissen (2007, pp. 17–18).

[70] See for example Oishi (2006).

With regard to question of epistemological assumptions it is to be noted that, in recent decades, *realism* has become a much more common position among both philosophers and working scientists, psychologists included.[71] Nevertheless, it is important to appreciate with regard to realism, that:

> Saying that something exists is not the same as saying that we know everything about it, nor that it has successfully been measured, nor that it is measurable in principle.[72]

So while many psychologists have indeed turned (sometimes explicitly, although usually not) from nominalism to realism, they rarely if ever adhere to *strong realism* (i.e. to holding the assumption that our familiar measurement methods correspond correctly to the true value of a specified quantity that exists independently of how we measure it). For instance, as the IRT scholars Borsboom and Mellenbergh recently stated:

> Psychometric models hypothesize the existence of latent variables to explain observed relations. If one has substantial confidence in the adequacy of the formulated psychometric model, one may estimate people's positions on the latent trait on the basis of observed scores. Note that the fact that, in doing this, the term 'estimation' is more appropriate than the term 'measurement'…[73]

Borsboom and Mellenbergh make a strict distinction between classical test theory, which is in its essence uncommitted to a realistic conception of *constructs*, and IRT, which indeed assumes the existence of *latent traits* (regardless of the measurement activity) about which it advances "testable hypotheses" with the aspiration of closing in on reality.[74] IRT is a version of a *weak* (or modest) *realism*. Psychometricians, after all use "models to help us go from the data we have, to the data we wish we had".[75]

As we have seen, the problem of IPCS was presented by economists as a question about accuracy and not about precision. Psychologists, however, had their own way of dealing with the issue, and were indeed extensively engaged with measurement problems connected to the IPCS problem. As a consequence, pragmatic solutions were established. Nevertheless, psychologists were, to use Harsanyi's terminology, extensively engaged with "the *psychological* question" of IPCS and not with the "metaphysical question." In fact, psychology and psychometrics could do very well without solving the metaphysical problems, which, more often than not, is how science in general progresses.

When we turn to happiness studies, however, and to the leap from collecting data to using and implementing the findings *outside of* psychological science (e.g. in public policy), the pragmatic approach to IPCS can no longer be taken for a substantial solution to IPCS. This is so on two levels. Firstly, even the most developed and well-established test cannot guarantee the representation of the same two real magnitudes. Secondly, the common tests, those that are actually used in broad range surveys, are not necessarily the most well-preformed and most valid ones. This is due to considerations such as cost and

---

[71] Maul (2013, p. 753). While some psychologists are aware of this shift and have explicitly suggested abandoning Stevens's basic definition of measurement, many other psychologists embrace Stevens's definition (usually without acknowledging the nominalist position it embodies). See Bond and Fox (2015, pp. 1–5), Michell (1997, 1999), Borsboom (2008, pp. 47–50).

[72] Maul (2013, pp. 753–754).

[73] Borsboom and Mellenbergh (2004, p. 118). See also Barsboom (2005), and Borsboom (2008) for an elaborated discussion of "latent variables".

[74] Barsboom (2005), chapters 2, 3.

[75] Zumbo (2007, p. 73).

convenience. Thus, for instance, using the one item test, which is not the most well-performed test for life-satisfaction, is common (see footnote 3). In light of these two problems we are forced to conclude that, while the practical solution to IPCS is well and good from a perspective *within* the science of psychology, when the jump is made to the world of public policy its limitations must be explicitly addressed.

3 The Ethical Approach

The problem of IPCS was presented in economics in the context of resource *allocation* and *redistribution*, in which the question of comparison between the utilities of different individuals is supposed to guide us with regard to their share of the redistribution (as part of a big zero-sum game). The original psychological research into satisfaction flowed from quite different concerns and objectives. This may partly explain why, from the start, the two scientific communities adopted different approaches—namely, the skeptical approach in economics, and pragmatic approach in psychology.

But our current context, in which public policy issues have come to the fore, establishes new challenges for both psychologists and economists who wish to see the practical implementation of their findings. This leap from theory to practice engenders various difficulties.[76] Our concern here is with the fact that when redistribution is made some real people gain at the expense of other real people. Redistribution policies that are based on measures of life satisfaction and other proxies for happiness are actually reshaping the *real* distribution of happiness among people (in case we accept a realistic point of view, of course). In this situation, one who implements the data by way of public policy cannot be satisfied with the pragmatic approach to the IPCS problem, as with psychometric methodology, nor is it sufficient to argue, with Harsanyi, that IPCS are logical *in principle*.

The results of a policy based upon life satisfaction data, all the above scientists would agree, is in any case unknown as far as the *real* satisfaction of people is concerned. As Harsanyi pointed out, the metaphysical question is inaccessible (with the exception, of course, of *Pareto improvement* cases, in which all are better off).[77] Pragmatic approaches, used at present by both psychologists and economists, therefore lack a necessary component, at least to the extent that redistribution policies are based upon them. One way out of this dilemma is to embrace a conception of IPCS that is *explicitly* based on an ethical-normative presumption – a presumption that may be accepted or rejected, but in any case is revealed to all.

Suggestions ascribing IPCS with value judgements (as a part of the solution) are not at all new and can be traced back to the 1970s. In the context of the debate over social indicators, for instance, in order to establish the possibility of implementing interpersonal welfare comparisons in redistribution decisions, Julian L. Simon (1974) suggested:

> The method is to choose measurable proxies for individual welfare… This method requires value judgements in the choices of proxy and of aggregation algorithm, and reasonable men may certainly differ on those value choices. But if some agreement on proxies and aggregation method can be obtained… then the methods suggested here should improve the process…[78]

---

[76] For specific problematic implications for public policy see Duncan (2013), Frey and Stutzer (2012).

[77] Indeed, it might be the case that part of the research is actually restricted to this kind of policy (or very close to it), but most decisions in public policy (especially the more interesting ones) are not such as this.

[78] Simon (1974, p. 66).

At least two value judgements are involved in every procedure of social welfare measuring, according to Simon: one is the choice of a proxy, the other is a choice of aggregation algorithm (the first and second aspects referred to in Sect. 1.2).

A similar recent illustration of the ethical-normative approach can be found in the thorough survey by Fleurbaey and Hammond (2004), in which a basic interpretation is made of the many ways of coping with the problem of IPCS with the aid of different kinds of proxies for measuring satisfaction/utility. Thus the basic procedure ascribed to social decision-making is a four-step procedure of the general form[79]:

1. (SWFL = Social welfare function) Specify an SWFL (f).
2. (CIG) Formulate a concept of individual good (Ui) that is appropriate as an argument for f.
3. (OP) Choose an observable proxy for each individual's Ui that is rich enough for these proxies to determine f(Un).
4. (D) Collect the necessary data about the proxy for each individual's Ui in order to determine f.

Following their construction of this procedure an argument is made by the authors that: Only the last step (D) involves exclusively factual propositions about the real world.[80]

Their point is that the process involves *normative judgments* through and through. In particular, step 3 is normative by its very nature, since "it is definitely a normative step to decide whether a given indicator is acceptable or not."[81]

Building on this framework I would like to suggest an interpretation that pushes the normative approach even further in the case of IPCS in happiness studies.

First, where exactly is the IPCS being dealt with in the case of the life-satisfaction-scale proxy? It should be noticed that, in the case of using life-satisfaction surveys, the *third step* of choosing the proxy inherently includes the process of normalizing by setting the ratings by a common scale (as explained in Sect. 1.2). So in this case the IPCS, or the trick of coming with IPCS by switching to a proxy for (Un), is made within this step. The first part of the fourth step, in this case, can be interpreted as collecting data that already consists of overcoming the IPCS problem. Now, in this very step, I suggest *two* distinct normative decisions are implicitly made (which one can either accept or reject): the normative (but not necessarily ethical-normative) decision of accepting the presumption that individuals' life satisfaction is to be represented by their self-reports[82]; and the *ethical*-normative decision of addressing the reports *as comparable* (made within the third step by normalizing the individual scales). This implicit decision/presumption is *ethical-normative* in the sense that it constitutes a judgement about ascribing similar meaning to scores of different people, or, in other words, it is the ethical conviction that similar scores of different people *deserve* to be regarded the same.

---

[79] Fleurbaey and Hammond (2004, p. 48). This description can be framed also with the first-aspect and second-aspect terminology (presented in Sect. 1.2), with the second and third steps together with the first part of the fourth step generating the first-aspect.

[80] Ibid.

[81] In an earlier prominent paper by Hammond (1991) a related explicit argument was raised to establish IPCS with the aid of: the values of the ethical observer, as influenced by that observer's understanding of the individual's psychology, and the observer's view of how society benefits from creating that individual or changing the individual's situation. Hammond (1991, p. 234).

[82] As suggested in the general case by Fleurbaey and Hammond.

4 Conclusion

As described in this paper, in economics, as in psychology, creative pragmatic approaches were developed during the second half of the twentieth century in order to cope with the obstacle of IPCS. This obstacle attracted particular attention in a scientific and philosophical context that has long since passed, namely the era when logical positivism ruled. Nevertheless, and as suggested here, in the current context of happiness studies and life-satisfaction data, when the findings of researches are implemented in public policy and as such help reshape economic distribution, it again becomes imperative to reconsider the exact implication of the pragmatic approach as a solution to IPCS.

A constructive suggestion that arises out of the discussion of this paper is that the policy maker who desires to implement data in public policy should accept the view that, because of redistribution implications, the IPCS problem may be coped with in so far as it is explicitly given an ethical-normative basis all the way to the bottom, as explained.

The costs of accepting such a solution are not, after all, such a high price to pay. For, in the first place, the procedure of computing averages, adding up the numbers, etc., in any case involves normative-ethical judgements (as explained in Sect. 1.2). Furthermore, other proxies of well-being, such as economic indicators, that are employed in the making of public policy decisions, also rest upon ethical-normative presumptions, as shown in Fleurbaey and Hammond (2004).

This ethical approach, which builds upon suggestions derived from the history of the methodology of economics, should be combined with that psychological methodology designed to make the surveys more and more precise and valid. The result would be the coming together of psychologists, economists and policy makers in the demonstration of a more accurate understanding of the strengths and limitations inherent in the implementation of Life Satisfaction data into public policy.

References

Adler, M. D. (2012). Happiness surveys and public policy: What's the use. *Duke LJ, 62,* 1509.

Adler, M. D. (2014). Extended preferences and interpersonal comparisons: A new account. *Economics and Philosophy, 30*(02), 123–162.

Angner, E. (2009a). Subjective measurement of well-being: Philosophical perspective. In H. Kincaid & D. Ross (Eds.), *The Oxford handbook of philosophy of economics*. London: Oxford University Press.

Angner, E. (2009b). The politics of happiness: Subjective vs. economic measures as measures of social well-being. In L. Bortolotti (Ed.), *Philosophy and happiness*. Basingstoke: Palgrave Macmillan.

Angner, E. (2011). The evolution of eupathics: The historical roots of subjective measures of well-being. *International Journal of Wellbeing, 1*(1), 4–41.

Audi, R. (Ed.). (2015). *The Cambridge dictionary of philosophy*. Cambridge: Cambridge University Press.

Barsboom, D. (2005). *Measuring the mind*. Cambridge: Cambridge University Press.

Bock, R. D., & Lieberman, M. (1970). Fitting a response model forn dichotomously scored items. *Psychometrika, 35*(2), 179–197.

Bond, T., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences*. London: Routledge.

Borsboom, D. (2008). Latent variable theory. *Measurement, 6,* 25–53.

Borsboom, D., & Mellenbergh, G. J. (2004). Why psychometrics is not pathological a comment on Michell. *Theory and Psychology, 14*(1), 105–120.

Broome, J. (1991). Utility. *Economics and Philosophy, 7*(01), 1–12.

Camerer, C. F., Loewenstein, G., & Prelec, D. (2004). Neuroeconomics: Why economics needs brains. *The Scandinavian Journal of Economics, 106*(3), 555–579.

Chang, H., & Cartwright, N. (2013). Measurement. In M. Curd & S. Psillos (Eds.), *The Routledge companion to philosophy of science* (pp. 411–420). Routledge.

Colman, A. M. (2008). *Oxford dictionary of psychology*. London: Oxford University Press.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*(4), 281–302.

Davis, J. (1992). The interpretation of interpersonal utility comparisons: Positive, normative or descriptive. *Journal of Income Distribution*, *1*(1), 73–90.

Diener, E. (1994). Assessing subjective well-being: Progress and opportunities. *Social Indicators Research, 31*(2), 103–157.

Diener, E. (2009). *Well-being for public policy*. London: Oxford University Press.

Diener, E. D., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment, 49*(1), 71–75.

Diener, E. D., & Seligman Martin, E. P. (2004). Beyond money: Toward an economy of well-being. *Psychological Science in the Public Interest, 5*(1), 1–31.

Diener, E., & Suh, E. M. (2000). *Culture and subjective well-being*. Cambridge, MA: MIT press.

Duncan, G. (2013). *Should happiness-maximization be the goal of government? The exploration of happiness* (pp. 303–320). Dordrecht: Springer.

Edgeworth, F. Y. (1877). *New and old methods of ethics*. London: James Parker and Co.

Edgeworth, F. Y. (1879). The Hedonical calculus. *Mind, 4,* 15.

Fleurbaey, M., & Blanchet, D. (2013). *Beyond GDP: Measuring welfare and assessing sustainability*. London: Oxford University Press.

Fleurbaey, M., & Hammond, P. J. (2004). Interpersonally comparable utility. In S. Barbera, P. J. Hammond & C. Seidl (Eds.), *Handbook of utility theory* (pp. 1179–1285). New York: Springer.

Frey, B. S., & Stutzer, A. (2009). Should national happiness be maximized. In A. K. Dutt & B. Radcliff (Eds.), *Happiness economics and politics: Towards a multi-disciplinary approach*. Cheltenham: Edward Elgar.

Frey, B. S., & Stutzer, A. (2012). The use of happiness research for public policy. *Social Choice and Welfare, 38*(4), 659–674.

Fumagalli, R. (2013). The futile search for true utility. *Economics and Philosophy, 29*(03), 325–347.

Hammond, P. J. (1991). Interpersonal comparisons of utility: Why and how they are and should be made. In J. Elster & J. E. Roemer (Eds.), *Interpersonal comparisons of well-being* (pp. 200–254). Cambridge: Cambridge University Press.

Hands, D. W. (2009). Effective tension in Robbins' economic methodology. *Economica, 76*(s1), 831–844.

Hands, D. W. (2012). The positive-normative dichotomy and economics. In U. Maki (Ed.), *Philosophy of Economics* (pp. 219–240). Elsevier.

Harsanyi, J. C. (1955). Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *The Journal of Political Economy, 63*(4), 309–321.

Harsanyi, J. C. (1982). Rational behavior and bargaining equilibrium in games and social situations. In A. Sen, B. Williams, & B. A. O. Williams (Eds.), *Utilitarianism and beyond* (pp. 39–62). Cambridge: Cambridge University Press.

Hausman, D. M., & McPherson, M. S. (2009). Preference satisfaction and welfare economics. *Economics and Philosophy, 25*(01), 1–25.

Isbell, J. R. (1959). Absolute Games. *Contributions to the Theory of Games*, *4*, 357–396.

Jevons, W. S. (1888). *Theory of political economy* (3rd ed.). London: Macmillan.

Jones, L. V., & Thissen, D. (2007). A history and overview of psychometrics. In C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics* (Vol. 26). Amsterdam: Elsevier B.V.

Kahneman, D., Diener, E., & Schwarz, N. (Eds.). (1999). *Well-being: Foundations of hedonic psychology*. New York City: Russell Sage Foundation.

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

Kahneman, D., Krueger, A. B., Schkade, D., Schwarz, N., & Stone, A. (2004). Toward national well-being accounts. *American Economic Review*, *94*(2), 429–434.

Kahneman, D., Wakker, P. P., & Sarin, R. (1997). Back to Bentham? Explorations of experienced utility. *The Quarterly Journal of Economics*, *112*, 375–405.

Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking*. New York: Springer.

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* (pp. 1–52). New York: Springer.

Layard, R. (2005). *Lessons from a new science*. New York: Penguin.

Layard, R. (2006). Happiness and public policy: A challenge to the profession. *The Economic Journal, 116*(510), c24–c33.

Little, I. M. D. (1950). *A critique of welfare economics*. London: Oxford University Press.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

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

Maul, A. (2013). On the ontology of psychological attributes. *Theory and Psychology, 23*(6), 752–769.

McCloskey, D. (2012). Happyism: The creepy new economics of pleasure. *The New Republic, 243.10*, 16–23.

Michell, J. (1997). Quantitative science and the definition of measurement in psychology. *British Journal of Psychology, 88*(3), 355–383.

Michell, J. (1999). *Measurement in psychology: A critical history of a methodological concept* (Vol. 53). Cambridge: Cambridge University Press.

Ng, Y. K. (1975). Bentham or Bergson? Finite sensibility, utility functions and social welfare functions. *The Review of Economic Studies, 42*(4), 545–569.

Ng, Y. K. (1982). Beyond Pareto optimality: The necessity of interpersonal cardinal utilities in distributional judgements and social choice. *Journal of Economics, 42*(3), 207–233.

Ng, Y. K. (1996). Happiness surveys: Some comparability issues and an exploratory survey based on just perceivable increments. *Social Indicators Research, 38*(1), 1–27.

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

O'Connor, B. P., Crawford, M. R., & Holder, M. D. (2015). An item response theory analysis of the subjective happiness scale. *Social Indicators Research, 124*(1), 249–258.

Oishi, S. (2006). The concept of life satisfaction across cultures: An IRT analysis. *Journal of Research in Personality, 40*(4), 411–423.

Oswald, A. J. (1997). Happiness and economic performance. *The Economic Journal, 107*(445), 1815–1831.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Denmarks Paedagogiske Institut, Copenhagen. Republished in 1980 by the University of Chicago Press.

Rasch, G. (1961). On general laws and the meaning of measurement in psychology. *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability* (pp. 321–333). Berkeley, CA: University of California Press.

Rasch, G. (1966). An item analysis which takes individual differences into account. *British Journal of Mathematical and Statistical Psychology, 19,* 49–57.

Rasch, G. (1977). On specific objectivity: An attempt at formalizing the request for generality and validity of scientific statements. In M. Blegvad (Ed.), *The Danish Yearbook of Philosophy*. Copenhagen: Munksgaard.

Reise, S. P., Ainsworth, A. T., & Haviland, M. G. (2005). Item response theory fundamentals, applications, and promise in psychological research. *Current Directions in Psychological Science, 14*(2), 95–101.

Robbins, L. (1934). *An essay on the nature and significance of economic sciene*. London: Macmillan & Co.

Robbins, L. (1997). *Interpersonal comparisons of utility: A comment* (pp. 199–204). Basingstoke: Palgrave Macmillan.

Rosenbaum, E. F. (1995). Interpersonal comparisons of utility: Positive, normative or value-laden? *Journal of Economic Methodology, 2*(2), 239–258.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*.

Schick, F. (1971). Beyond utilitarianism. *The Journal of Philosophy, 68*(20), 657–666.

Sen, A. (1979). Interpersonal comparisons of welfare. In M. J. Boskin (Ed.), *Economics and human welfare: Essays in honor of tibor scitovsky* (pp. 183–201). Academic Press.

Simon, J. L. (1974). Interpersonal welfare comparisons can be made and used for redistribution decisions. *Kyklos, 27*(1), 63–98.

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

Thurstone, L. L. (1931). *The reliability and validity of tests*. Ann Arbor, MI: Edwards Brothers.

Veenhoven, R. (2010). Greater happiness for a greater number. *Journal of Happiness Studies, 11*(5), 605–629.

Waldner, I. (1972). The empirical meaningfulness of interpersonal utility comparisons. *The Journal of Philosophy, 69*, 87–103.

Zumbo, B. D. (2007). Validity: Foundational issues and statistical methodology. In C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics* (Vol. 26). Amsterdam: Elsevier B.V.