



The evolution of eupathics: The historical roots of subjective measures of wellbeing

Erik Angner

Abstract: This paper traces the historical roots of subjective measures of wellbeing, that is, measures designed to represent happiness, satisfaction, or other “positive” or desirable mental states. While it is often suggested that these measures are a modern invention, I argue that they have a long and rich history that conforms to Theodore M. Porter’s general account of measurement in social and behavioral science. Subjective measures emerged in marital success studies, educational psychology, and personality psychology in the 1920s and 30s, and were further shaped by the epidemiology of mental health, gerontology, and the social indicator movement in the 1960s and 70s. Consistent with Porter’s account, these measures emerged in applied rather than theoretical branches of social and behavioral science, and they did so not as a result of physics envy, but rather as a result of a moral impulse to improve society; quantification was intended to make up for perceived deficiencies in unaided human judgment; and radical disagreements about the nature of wellbeing did not impede efforts to measure it – indeed, in time, there was considerably more agreement about how to measure wellbeing than about how to define it.

Keywords: eupathics, happiness, measurement, satisfaction, wellbeing, well-being, positive psychology

1. Introduction

It is often suggested that subjective measures of wellbeing – measures designed to represent happiness, satisfaction, or other “positive” or desirable mental states – are a relatively new invention. Bruno S. Frey and Alois Stutzer, for example, write: “Recently, great progress has been achieved in economics: happiness has been seriously measured, and many of its determinants have been identified” (Frey & Stutzer, 2000, p. 145). Two years later, the same authors note that economists traditionally have given little attention to questions of happiness, and add: “In the past few years the situation has changed: A number of economists see an advantage in measuring subjective well-being as expressed by individuals themselves” (2002, p. vii). Similarly, it is often suggested that the so-called science of happiness – the body of empirical generalizations developed using subjective measures – constitutes a new development. This suggestion is evident, e.g., in the subtitle of Richard Layard’s book *Happiness: Lessons from a new science* (2005), and it is supported by authors like Martin E. P. Seligman, who grumbles: “For the last half century psychology has been consumed with a single topic only – mental illness” (Seligman, 2002, p. ix).

The impression that the systematic study of happiness and the like constitutes a recent development is partially due to its association with the positive psychology movement (cf.



Seligman & Csikszentmihalyi, 2000). This movement, which emerged in the 1990s in large part due to Seligman's efforts, is motivated by the belief that traditional psychology has spent an inordinate amount of time examining psychological pathology. According to positive psychologists, traditional psychology overlooked "positive" mental states because they were seen as derivative from or less authentic than negative ones, and therefore less worthy of study; in contrast, positive psychology proceeds from the assumption that positive mental states are no more derivative and no less authentic than negative ones, and consequently worthy of attention in their own right (Seligman, 2002; Seligman & Csikszentmihalyi, 2000). Because of their implicit or explicit association with positive psychology, which as a self-conscious subdiscipline of psychology is a new development, it may seem as though subjective measures of wellbeing are new too.

The goal of this paper is to trace the historical roots of measures of happiness, satisfaction, and the like. My thesis is that such measures, far from being a novel invention, have a long and rich history, and that this history conforms to the general account of measurement in social and behavioral science offered by Theodore M. Porter (1995). The appearance of a science of happiness was perhaps marked by a 1917 article promoting a discipline of "eupathics," defined as the study of "the *well being of the normal*" (Myerson, 1917, p. 344). Modern efforts to develop measures of happiness, etc., are part of an uninterrupted research stream going back at least to the 1920s and 1930s. Subjective measures emerged in marital success studies, educational psychology, and personality psychology in the 1920s and 30s, as tools developed for the study of personality were applied to happiness and satisfaction; the measures were further shaped by the epidemiology of mental health, gerontology, and the social indicator movement in the 1960s and 70s, as measures of happiness and satisfaction were recruited as proxies for mental health and wellbeing in large, representative samples. Consistent with Porter's account, subjective measures of wellbeing emerged in applied rather than theoretical branches of social science, where they were developed not as a result of physics envy but of a moral impulse to improve society; quantification was intended to make up for perceived deficiencies in unaided human judgment; and radical disagreements about the nature of wellbeing did not impede efforts to measure it – indeed, in time, there was considerably more agreement about how to measure wellbeing than about how to define it.

As an exercise in what Hasok Chang (2004) has called history of science as *complementary science* – that is, the "continuation of science by other means" (p. 249) – the history of subjective measures of wellbeing should be of interest not just to historians of psychology, but to practicing social and behavioral scientists as well as to anyone interested in the application of the science of happiness to public policy. As the quotes in the first paragraph above indicate, even highly prominent researchers of happiness and satisfaction appear unaware of the longevity and richness of the research stream of which they are part. The story offers further evidence for Chang's thesis that history can serve to recover knowledge – about natural and social phenomena as well as about previous efforts to measure them – that appears to have been forgotten; to develop critical awareness of contemporary efforts, their prospects and limitations; and to suggest novel experimental and theoretical developments.

2. The emergence and evolution of subjective measures

In this section, I outline how measures of happiness and satisfaction emerged in educational psychology, marital success studies, and personality psychology in the 1920s and 30s, and how those measures were further shaped by the epidemiology of mental health, gerontology, and the social indicator movement in the 1960s and 70s (see Appendix A for a timeline). As my

starting point, I take the references in a prominent contemporary review article (Diener, Suh, Lucas, & Smith, 1999) as well as the sources quoted therein; for the early studies, I have relied heavily on the review article by Warner Wilson (1967). I group the studies by the discipline in which they appeared in order to emphasize the broad movements that generated and sustained an interest in happiness, and the like, as well as in the methods required to measure such things. This is not to suggest that work in different disciplines was entirely independent: in fact, as we will see, there was a fair amount of cross-pollination. Furthermore, I do not mean to suggest that the historical sketch below is in any interesting sense complete: given how rich the history turns out to be, and given my focus on broad trends, I can offer only brief sketches of individual figures, projects, measures, and methods, and hint at the cultural, intellectual, academic, social, and economic context of which they were part.

2.1 Marital success studies

Two of the earliest studies on happiness were penologist and social worker Katharine Bement Davis's *Factors in the sex life of twenty-two hundred women* (1929) and psychiatrist G. V. Hamilton's *A research in marriage* (1929). Davis, who received her Ph.D. in economics, was interested in gathering "adequate data as to both the physical and mental facts of the sex life of the normal individual" (Davis, 1929, p. ix). She also wanted to explore correlations between facts about women's sex lives and other aspects of their lives. Thus, one of the questions given to the roughly 2,200 subjects was the following: "Do you consider your life on the whole (a) happy, satisfactory, successful; (b) unhappy, unsatisfactory, unsuccessful? In each case why?" (p. 89). Hamilton's study was similar in spirit, except that his subjects were married couples. He gave a window into what motivated these studies when he wrote:

My standpoint is that of the psychiatrist who believes that subjective phenomena, as these are experienced by the persons who report their occurrence, do not need to be translated into anything else in order to be dealt with as objectively as we deal with all other biological phenomena (Hamilton, 1929, p. xi).

Specifically, Hamilton wanted to know how satisfied his subjects were with their marriages. To this end, he used responses to a set of questions about satisfaction to compute a "satisfaction-grade," which he also referred to as an "index of spousal satisfaction/dissatisfaction" (pp. 78-79). The satisfaction grade was thought to represent what Hamilton called "marital success" (p. 8).

The work of Davis and Hamilton clearly inspired Lewis M. Terman, Stanford professor of education and psychology, proponent of large-scale standardized testing, and famous above all for the Stanford-Binet intelligence test (Hilgard, 1957). In a 1938 book called *Psychological factors in marital happiness*, Terman described his project as follows: "We have selected as the theme of our study that aspect of the successful marriage which may be designated as marital happiness, and we wish to ascertain, if possible, what psychological factors are demonstrably associated with this state" (Terman, 1938, p. 2). To this end, Terman distributed questionnaires to 792 couples from the "middle and upper-middle classes of urban and semiurban Californians" (p. 13). He constructed a "marital happiness score" on the basis of "(1) subjective ratings of the happiness of the marriage; and (2) factual information on husband-wife agreement or disagreement about various matters," in which the subjective rating "was allowed a heavy weighting" (p. 3). For the subjective rating, respondents were given the question "Everything considered, how happy has your marriage been?" and asked to "Draw a circle around 1, 2, 3, 4,

5, 6, or 7," where the options were "1 = Extraordinarily happy," "2 = Decidedly more happy than the average," "3 = Somewhat more happy than the average," and so on (p. 440).

A decade after Davis (1929), but drawing above all on Terman (1938), sociologists Ernest W. Burgess and Leonard S. Cottrell published a study called *Predicting success or failure in marriage* (1939). Their study had three goals: to define "marriage adjustment," to identify what factors were associated with marital success or failure, and to determine whether it was possible to predict ahead of time which marriages would lead to happiness and which would not (Burgess & Cottrell, 1939, p. 15). Burgess and Cottrell motivated their study of marital success by arguing that marital adjustment and incompatibility had become a social problem and therefore of public concern (p. 1). Defining their terms, the authors wrote:

A well-adjusted marriage from the point of view of this study may then be defined as a marriage in which the attitudes and acts of each of the partners produce an environment which is favorable to the functioning of the personality of each, particularly in the sphere of primary relationships (Burgess & Cottrell, 1939, p. 10).

In their study, 526 couples were given the following prompt: "Appraisal of marriage: very unhappy.....; unhappy.....; average.....; happy.....; very happy....." and asked to check the relevant box (Burgess & Cottrell, 1939, p. 422). In practice, the authors used happiness as the criterion by which they judged both adjustment and success (p. 30).

In sum, Davis, Hamilton and their followers were interested in a cluster of concepts – including "happiness," "adjustment," "success," and "satisfaction" – which by and large were treated as interchangeable, and in the association of those concepts with personality, background, and sexual factors. These authors clearly believed that "subjective phenomena" were worthy of study in their own right and that they could be explored directly – without being "translated into anything else," as Hamilton (1929, p. xi) put it – by means of straightforward questions about respondents' happiness or satisfaction. The description of inadequate marital adjustment as a social problem and of public concern underscores the fact that they thought the results of their studies could be helpful in alleviating social problems. Interestingly, some of these researchers appear to have been more concerned with the happiness or satisfaction of the marriage than with that of the people in it (Burgess & Cottrell, 1939, p. 422; Terman, 1938, p. 440). This literature might have been motivated not primarily by a desire to make people happy, but by a desire to ensure a sufficiently high population growth rate.

2.2 Educational psychology

Perhaps the most influential early study of happiness, however, was Goodwin Watson's "Happiness among adult students of education" (1930). Watson, who was a professor of education at Columbia University's Teachers College, introduced his topic in the following way:

No human quest may claim a larger following than that for happiness and satisfaction in life. Even the highest ethics tends to justify itself by its contribution to human happiness. The Beatitudes themselves constitute an attempt to formulate the conditions of the most blessed existence. Certainly any educational program in modern times is likely to be justified only in terms of its direct or indirect contribution to human happiness. It becomes, therefore, extraordinary almost beyond belief that so few attempts have been made to

apply the techniques of psychological study to the understanding of happiness (Watson, 1930, p. 79).

What Watson called “the highest ethics” might have referred to Aristotle’s notion of *eudaimonia* or to the utilitarians’ maxim of the greatest happiness for the greatest number, and “[the] Beatitudes” to the blessings said by Jesus in the Sermon on the Mount. Watson clearly disapproved of the fact that psychologists had not explored systematically how to achieve greater happiness – as instructed by philosophical and religious authorities – and he set out to make up for the deficiency.

Watson distributed anonymous questionnaires to 388 graduate students of education, and used a wide variety of measures to elicit degrees of happiness. As we will see in Section 3.4, some of the measures were remarkably sophisticated. Examining the responses, Watson noted that subjects in general “tended to think of themselves as above the average on each happiness item” (p. 84). Subjects who scored low to average in aggregate happiness “appeared to believe that they gave an impression of greater happiness than was justified,” whereas those who scored high “felt no sense of masquerade” (p. 87). Watson also studied the correlation between aggregate happiness and subjects’ personality traits and living conditions. He found that happiness was not predicted by IQ scores (p. 88), age, whether the subject came from a small town, physical defects, number of siblings, parents’ demand for obedience, having parents older than 40, whether the mother had a career, parents’ divorce or separation, school marks, extra-curricular activities, crushes, a wise sex education, or (perceived) popularity with the opposite sex, among other things (pp. 94-96). Perceived harmony between parents did predict happiness, however (p. 94), as did (among other things) a preference to spend a few hours with the most delightful known companion of the same sex over one of the opposite sex, an ability to give a lecture to high school students about sex, an ability to administer a large group of people, an absence of strong fears, health during adolescence, an ability to fuse well, an absence of shyness or timidity, and marriage (p. 97).

Another important early contribution to the literature on happiness was a thesis written by Randolph C. Sailer under Watson’s supervision. The thesis was published as *Happiness self-estimates of young men* (Sailer, 1931). While Sailer largely followed the approach taken by Watson (1930), he was also inspired by Hamilton (1929) and others (Sailer, 1931, p. 7). This fact indicates that there was cross-pollination at an early stage: researchers with different backgrounds were not ignorant of their predecessors’ efforts and in fact built upon them. Distributing questionnaires to 500 young men across the United States, Sailer found among other things an association between happiness on the one hand and religion as well as sociability on the other (p. 99). He added: “An especially notable relationship, the largest found in the entire study, connected those reporting themselves ‘more even, on a level’ rather than ‘more up-and-down, now happy, now depressed’ with the happier groups” (pp. 98-99).

In the mid-1930s, a number of studies followed Watson and Sailer in exploring the relationship between various aspects of personality and happiness scores. Many of these studies found a relationship that turned out to be surprisingly weak. Thus, George W. Hartmann’s 1934 study “Personality traits associated with variations in happiness” found little correlation between personality traits and experienced happiness (Hartmann, 1934, p. 211). Hartmann, who was a psychologist at the Pennsylvania State College, used the term *felicity* instead of happiness (p. 209). Similarly, Percival M. Symonds’s “Happiness as related to problems and interests” (1937) concludes: “Happy and unhappy [people] are remarkably alike in their problems and interests. The unhappy do not have peculiar problems but make less

satisfactory adjustments to their problems” (Symonds, 1937, p. 293). Like Watson and Sailer, Hartmann and Symonds both had ties to the Teachers College at Columbia University.

In sum, educational psychologists like Watson, Sailer, Hartmann, and Symonds agreed with Davis, Hamilton, and their followers that subjective phenomena like happiness and satisfaction were worthy of study in their own right. In contrast to some students of marital success, Watson et al. were unambiguously interested in happiness, satisfaction, and felicity – which they appear to have treated as interchangeable – as a property of the individual person rather than his or her marriage. Unlike the authors who preceded them, Watson and his followers used a wide variety of sometimes remarkably sophisticated measures to assess the degree of happiness exhibited by their subjects. As we will see in Section 3.4, the fact that they used a variety of measures allowed them to compare the performance of different measures and to argue that simple self-report measures were both reliable and valid and could often substitute for more extensive batteries of questions. That line of work set the tone for much of the research that followed.

2.3 Personality psychology

Though efforts to measure happiness and satisfaction have roots in marital success studies and educational psychology, the discussion so far also implicates the discipline of personality psychology. According to a modern definition: “Personality psychology is the scientific study of the whole person. The goal of personality inquiry is to provide a scientifically credible account of human individuality” (McAdams, 2001, p. 11308). At the time Davis, Watson, and their followers started studying happiness, personality psychology was in the process of emerging as an independent subdiscipline of psychology. While the effort to develop reliable methods for studying differences in personality traits across people and over time was not completely new, the “1921-1938 period was a time of intense research activity,” by the end of which the field of personality psychology was officially established (Winter & Barenbaum, 1999, pp. 8-9). The event is marked by the appearance of the first authoritative textbook (Allport, 1937) and a landmark study (Murray, 1938) (cf. McAdams, 2001, p. 11309). Interestingly, Murray went on to direct the work of Herbert Jeremy Goldings at the Harvard Psychological Clinic, to whom we will return in the next section (cf. Goldings, 1954, p. 30). From its beginnings in the aftermath of World War I, 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). Hence:

Building on the pioneering work of Francis Galton and Alfred Binet on mental testing and spurred by the mobilization of large military forces in World War I, psychologists began to invent self-report, multi-item tests to assess individual differences in personality functioning (McAdams, 2001, p. 11309).

The hope was that a new exact science of personality would “furnish assistance to a corporate culture and a government suddenly confronted by dramatic changes and the need to ‘manage’ and control an American population that had suddenly become larger, more diverse, and ‘difficult’” (Winter & Barenbaum, 1999, p. 5).

Scientific and methodological developments relating to happiness and satisfaction during the late 1920s and 1930s are usefully seen against this background. The happiness research surveyed above was obviously driven by a desire to measure and account for personality characteristics across individuals. It was also apparently motivated by an ambition to enhance the happiness of people, or their marriages, and to help governments alleviate social problems.

Seeing the measurement of happiness as an outgrowth of personality psychology is useful for answering Watson's question at the very beginning of his 1930 paper (cited in Section 2.2). Watson asked why it had taken so long for psychologists to explore empirically how happiness is distributed, what factors predict it, and how one should go about promoting it. The link with personality psychology goes a long way toward answering this question: it was not until after Galton, Binet, and the experiences of World War I that psychologists developed the confidence that they could reliably measure personality characteristics like happiness.

In conclusion, the emergence of personality psychology as an independent subdiscipline of psychology was a critical step in the history of measures of happiness and satisfaction and the science of happiness. It was only with the establishment of personality psychology, thanks to people like Allport and Murray, that many psychologists developed the confidence that they could reliably and validly measure attributes of the person and that the results could be useful to governments and corporations. This hypothesis explains why, before the 1920s, few psychologists attempted to systematically measure happiness and satisfaction: the psychologists of the time had not yet convinced themselves that it was possible to measure attributes of the person. The hypothesis also explains why during the 1930s many psychologists thought that such a study was not only possible but long overdue. Given that for philosophical, religious, or other reasons they thought that happiness was important to understand and promote, the conviction that it was possible to measure happiness quickly produced a desire to measure it.

2.4 Synthesis and further development

After these developments, in short order, there emerged a number of studies that drew equally on the results from marital happiness studies and educational psychology. For example, Hornell Hart's inventive contribution *Chart for happiness* (1940) explicitly drew on both.¹ Hart, a professor of sociology at Duke University, asked: "One of the basic purposes of mankind is to be happy. Can recent advances in scientific thinking tell us more and more effectively what to do in order to be happy, and in order to help make our fellow human beings happy?" (p. 16). The book was written both for those who "sense the fact that they are not living on as full a tide of happiness as they might attain" and for "psychologists, sociologists, psychiatrists, physicians, educators, social workers, pastors" and other professionals who are concerned with the happiness of others (pp. v-vi). Hart paid more attention to the nature of happiness than his predecessors, offering not one but two "operational" definitions of the term: according to the one, people are happy insofar as they are prone to saying sincerely that they are happy; according to the other, people are happy insofar as they are in whatever state of consciousness that they seek to attain or maintain (pp. 182-184; cf. Section 3.3 below).

Hart was deeply impressed by measurement instruments used in other scientific disciplines and hoped to develop something equally useful for happiness. Having discussed intelligence tests, the fever thermometer, and the electrocardiograph, he went on to write:

Chart for Happiness presents and explains another such invention, the Euphorimeter, produced by the researches of various social scientists in various universities and colleges. This new instrument is to measure, not intelligence, or fever, or cardiac pulsations, but the happiness of those who submit themselves

¹ Specifically, Hart discussed data from Hamilton (1929), Watson (1930), Sailer (1931), Terman (1938), and Burgess and Cottrell (1939) (cf. Hart, 1940, pp. 173-174). Hart was well versed in the research that preceded him.

to it. It is to diagnose, not mental aptitudes, nor physical disease, but the causes of human anguish (p. v).

Hart wanted happiness scores to be measurable on an “objective scale ... which will have meaning for ordinary people” (p. 177). Hence, he normalized the scores and assigned a score of zero to the point that divides happiness from unhappiness and 100 to the average happiness of all persons tested (pp. 175-176). As Hart described it:

The scale of this instrument is somewhat like the scale of a centigrade thermometer. The zero on the Euphorimeter is the dividing point between happiness and unhappiness. From zero down, the scores mean deeper and deeper unhappiness. From zero up, the scores indicate greater and greater happiness. The 100-degree point is set by the average happiness of the general population (p. 19).

The unit of the Euphorimeter is called a *Euphor-unit*, and is defined as “one one-hundredth of the difference between the zero point and the average happiness score” (p. 178). Hart’s method for normalizing happiness scores allowed him to compare the results of various studies, as shown in Figure 1.

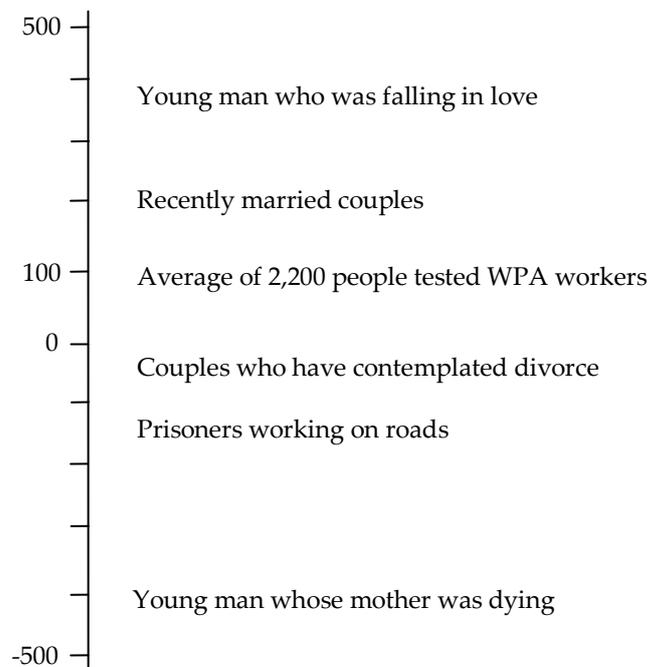


Figure 1. The Euphorimeter scale, with some typical scores.²

For purposes of his own research, Hart developed two separate Euphorimeters. Because people’s happiness fluctuates from moment to moment, Hart wrote, “it has been necessary to devise two types of Euphorimeters – one to measure happiness ‘at the moment,’ the other to measure as closely as possible the general level of happiness or unhappiness on which one lives in the long run” (p. 21). One advantage of the At-the-Moment Euphorimeter was that it allowed him to develop a dense record of participants’ happiness levels over time. Hart used this Euphorimeter to draw charts showing how the happiness of three subjects, one who was falling in love, one whose mother was dying, and one who wondered why not everybody committed suicide, evolved over time (see Figures 2, 3, & 4 below).

² Figures 1, 2, 3, & 4 are reproductions based on Hart (1940, pp. 20-24).

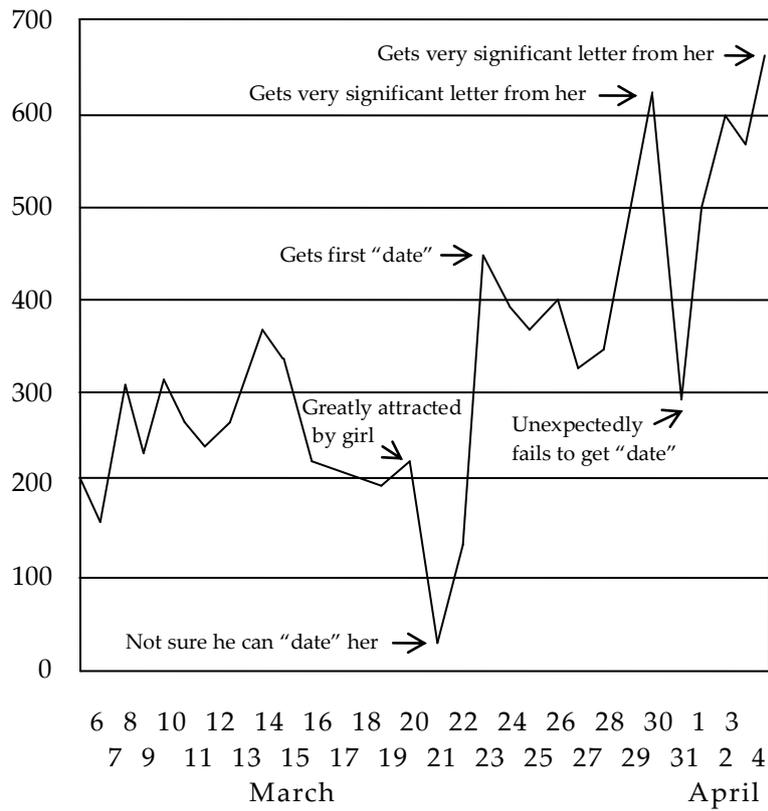


Figure 2. Happiness fluctuations of a student who was falling in love.

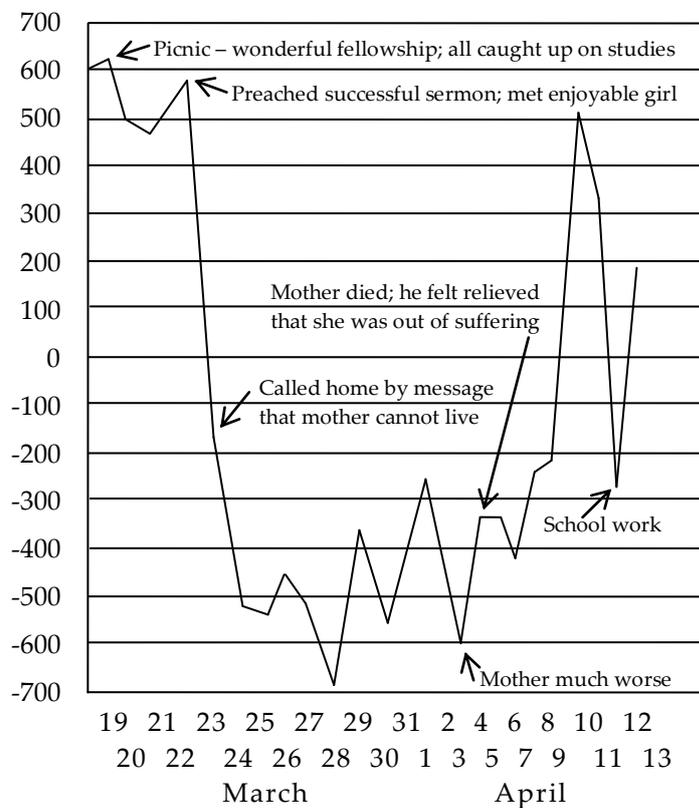


Figure 3. Happiness fluctuations of a student whose mother was dying.

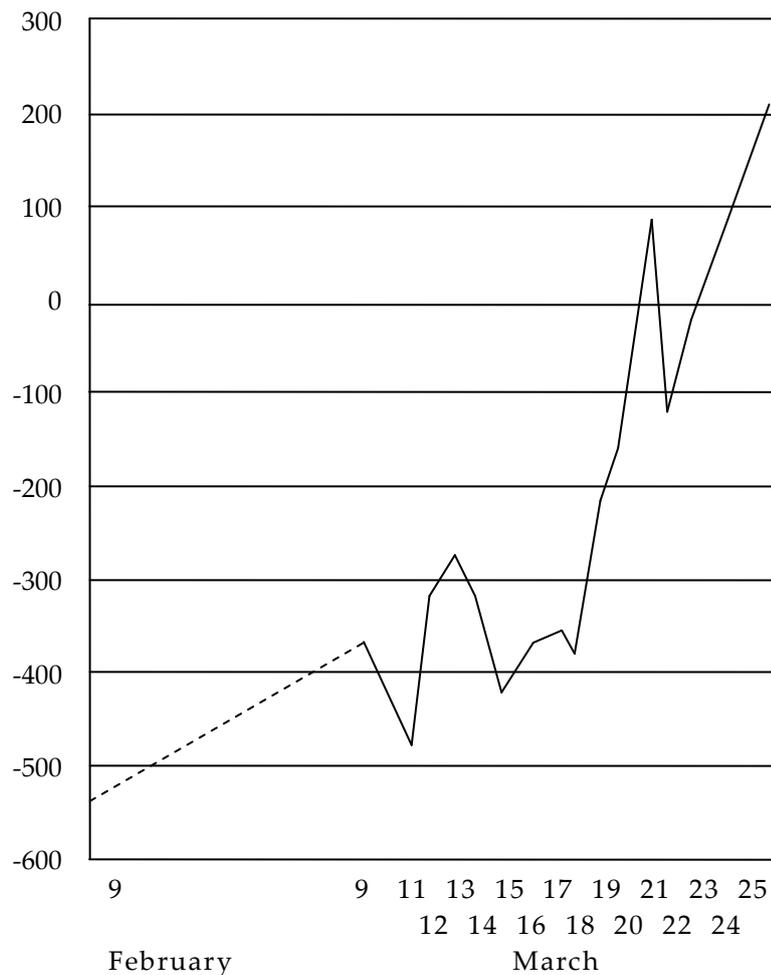


Figure 4. Happiness fluctuations of a college boy who wondered why everybody did not commit suicide.

Hart found the case of the suicidal college student particularly interesting. Taking the test apparently made the student realize that he was unusually unhappy and for no good reason; by tracking his Euphorimeter readings carefully, it took him no more than a month to rise from the depths of despair to above-average happiness. Hart concluded: "Perhaps you can make equally great progress, if you also are unhappy" (p. 25).

In 1954, under Murray's supervision, Herbert Jeremy Goldings of the Harvard Psychological Clinic published another study, which drew on research both on marital happiness and educational psychology. Goldings suggested that he wanted to tap into "a commercially vigorous popular concern with happiness (books, magazines, etc.) indicating, perhaps, that in American culture people are not happy, or not as happy as they want to be, or not as happy as they feel they should be" (Goldings, 1954, p. 30). Goldings's article is interesting in part because it contains an explicit discussion of "the intrinsic nature and characteristics" of happiness (p. 31). He defined happiness as a physiological state, though one that has consequences that are accessible through introspection. His article is also interesting because it uses new methods to measure happiness. Other than using measures adapted from Watson (1930) and Sailer (1931), Goldings asked his subjects to rate the happiness and satisfaction of 30 photographs of faces with ambiguous facial expressions, and proposed that

the result could serve as an “indirect” measure of happiness, on the theory that subjects would project their own happiness onto the ambiguous facial expressions (Goldings, 1954, p. 33; cf. Section 3.4 below). Using 20 Harvard undergraduates as subjects, Goldings confirmed Watson’s finding that subjects tend to rate their own happiness as greater than the average (Goldings, 1954, p. 36). Goldings also found “a highly significant degree of agreement among experienced judges’ ratings of happiness” and a high correlation between avowed happiness and judges’ ratings (p. 46). However, because it did not behave as anticipated, he ended up rejecting the indirect measure as a measure of happiness.

In brief, Hart and Goldings were well aware of their predecessors in marital success studies and educational psychology, and they sought ways to synthesize and build upon that research. Both believed that individual happiness was worthy of study in its own right, that the scientific approach to happiness had particular power and potential, that self-report measures of happiness were adequate for a wide variety of purposes, and that scientific studies could help people become happier. By comparison with their predecessors, both Hart and Goldings offered a more explicit discussion about the nature of happiness. They also continued the educational psychologists’ efforts to develop and test novel ways to assess happiness. Thus, Hart developed a procedure, inspired by the centigrade thermometer, for standardizing happiness scores. Hart also started tracking people’s happiness over time, which is particularly interesting in light of the affinities with the much more recent work by Daniel Kahneman and co-authors (e.g., Redelmeier & Kahneman, 1996). Meanwhile, Goldings developed a new projection-based indirect measure. Not all their efforts succeeded: because Goldings’s indirect measure failed to behave as expected, he abandoned the project of using such measures as measures of happiness. Both Hart and Goldings expressed a desire to reach a wider audience, and perhaps to tap into what Goldings (1954, p. 30) had called the “commercially vigorous popular concern with happiness.”

2.5 The epidemiology of mental health

In the late 1950s and early 60s a number of authors with a background in the epidemiology of mental health recruited measures of happiness and satisfaction for use as measures of mental health in large, sometimes representative samples. This move, as we will see shortly, appears to have had an enormous impact on the use and interpretation of measures of happiness and satisfaction.

A landmark study in the epidemiology of mental health is *Americans view their mental health*, by Gerald Gurin, Joseph Veroff, and Sheila Feld (1960). As the title indicates, Gurin and his co-authors were primarily interested in how individuals themselves – as opposed to mental health experts – saw their mental health (Gurin et al., 1960, pp. 3-4). The authors adopted a multiple-criterion perspective: they believed that mental health was best measured by using a battery of criteria, one of which was whether people felt happy. Thus, Gurin et al. did not identify happiness and mental health, but saw the former as an important component or indicator of the latter. One of the questions they asked was the following: “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?” (p. 411). Unfortunately, although they discussed what the subjects thought contributed to their happiness, Gurin et al. said little about what they meant by the term (pp. 22-28). Because they wanted a representative sample, Gurin et al. enrolled 2,460 randomly selected, non-institutionalized American adults (p. 3). In this sample, Gurin et al. found an association between unhappiness and pessimism and uncertainty about the future (p. 35), but no connection between the sex of the subject and his or her happiness (p. 42). They found that

older people were less happy than younger (p. 44), and that those with more education were happier than those with less (p. 46). They also found that happiness was positively related to income (p. 216).

Another important contribution to the epidemiology of mental health is Norman M. Bradburn and David Caplovitz's book *Reports on happiness: A pilot study of behavior related to mental health* (1965), which built on Gurin et al. (1960). According to Bradburn and Caplovitz, this study was "an effort to develop, for psychological and behavioral phenomena, time-series studies comparable to those that are commonplace in economics and demography" (Bradburn & Caplovitz, 1965, p. 1). As the following quote indicates, they were particularly interested in "psychological well-being":

Its long-range objective is to conduct periodic inventories of the psychological well-being of the nation's population. From such inventories it is possible to determine for the first time the extent to which feeling states of the population are affected by major social trends, national and local crises, and changes in the economic and social structures, as well as by patterned events in the life cycles of individuals (p. 1).

It appears that the authors identified "psychological well-being", "mental health", and "happiness" into a one-dimensional entity. In Bradburn and Caplovitz's words:

The underlying assumption of this research is that there is a dimension variously called mental health, subjective adjustment, happiness, or psychological well-being, and that individuals can be meaningfully described as being relatively high or low on such a dimension (p. 1).

The authors were comfortable using self-report measures, as they wrote: "The most fruitful starting point in developing an instrument to measure a dimension of mental health seemed to be people's own estimates of their level of psychological well-being or distress" (p. 5). Using the specific measure employed by Gurin et al. (1960), Bradburn and Caplovitz administered questionnaires to 2,006 members of 450 households in four rural Illinois communities (Bradburn & Caplovitz, 1965, pp. 3-5) and concluded among other things that happiness was positively correlated with education and income, negatively correlated with age, and uncorrelated with sex (p. 56).

In sum, mental health professionals like Gurin, Bradburn and Caplovitz were concerned with the causes, distribution, and promotion of mental health in the population, and decided that measures of happiness and satisfaction could be recruited for the purpose. They thought it legitimate to use a measure of happiness as a measure of mental health because they considered happiness either identical to, a component of, or an indicator of mental health or psychological wellbeing. To assess individual happiness, these authors were comfortable using a single straightforward question, answered on a three-point scale, which remains in use today. As compared to previous research, the effort of the mental health professionals was different in important respects. Above all, Gurin et al. wanted to know the distribution of happiness in the population as a whole, so they started using large, random, and therefore representative, samples from the national population; Bradburn and Caplovitz wanted to know not only how happiness was distributed in the population, but also how the distribution developed over time, so they started collecting time series data, much like Hart (1940) had. This approach would be picked up and further explored by representatives of the social indicators movement, to which we will return in Section 2.8. In recent years, of course, the effort to track the happiness and satisfaction of the nation over time has received a great deal of attention under

headings like “national indicators of subjective well-being” and “national well-being accounts” (e.g., Diener, 2006; Kahneman, Krueger, Schkade, Schwarz, & Stone, 2004).

2.6 Review and call for more research

In 1967, University of Alabama psychologist Warner Wilson published the first major review article of research on happiness, or as he said (borrowing a term of Goldings’s), “avowed happiness.” Wilson reviewed the research discussed above, his 1960 dissertation in the area, and more. On the basis of this meta-analysis, Wilson concluded that “avowed happiness can be determined reliably,” and that the “facts ... support the validity of self-ratings” (1967, pp. 294-295). On the basis of the studies he reviewed, Wilson inferred: “The happy person emerges as a young, healthy, well-educated, well-paid, extroverted, optimistic, worry-free, religious, married person with high self-esteem, high job morale, modest aspirations, of either sex and of a wide range of intelligence” (p. 294). Under “Suggestions for further research,” Wilson offered the following: “Comparability of results might be increased if all investigators used the same measure. The designations ‘very happy,’ ‘pretty happy,’ and ‘not too happy’ used by Gurin et al. (1960) and by Bradburn and Caplovitz (1965) seem as good as any” (Wilson, 1967, pp. 304-305).

Wilson ended his article by calling for more experimental research, in which the happiness of subjects is manipulated through therapy or other means. “Studies involving direct attempts to manipulate the well-being of individuals are most desirable,” he maintained (p. 305). He continued:

A story is told of how a small group of college men increased the poise and popularity, and presumably the happiness, of a female student by going out of their way to respond to her as though she were attractive. The possibility of systematically developing and applying such principles and techniques seems exciting indeed! (p. 305)

While nobody, to my knowledge, has performed this particular experiment under controlled conditions – presumably because of ethical constraints – more recent happiness scholars have started performing controlled studies examining the causes of happiness. Many of them are interested in developing interventions that can reliably be used to enhance people’s happiness (cf. Lyubomirsky, 2008).

2.7 Gerontology

Around the same time, a number of studies on happiness and satisfaction had appeared in the field of gerontology (Larson, 1978). According to David L. Adams (1971), a concern with “individual well-being” was one of the factors contributing to the development of gerontology in the first place (p. 64). He continued: “However, the difficulty of trying to assess ‘individual well-being’ has resulted in a variety of concepts, definitions, and measurements” (p. 64). The concepts included “morale,” “successful aging,” and “personal adjustment.” From the start, according to Bernice L. Neugarten, Robert J. Havighurst, and Sheldon S. Tobin (1961), there were two different approaches to measuring psychological wellbeing or wellbeing in gerontology. The one, according to Neugarten et al., focused upon “the overt behavior of the individual and utilizes social criteria of success or competence” (p. 134). This approach was criticized, however, for relying too heavily on the assumption that social competence and participation were closely associated with wellbeing, and for imposing on the subjects the value judgments of the investigators (p. 134). The other approach, which emerged at least in part as a

reaction to the first, relied instead on an “internal frame of reference” in which attributes like success and competence figured only indirectly. Hence: “Here the variables to be measured have been the individual’s own evaluations of his present or past life, his satisfaction, or his happiness” (p. 134). In this approach:

The assumptions are, whether or not explicitly stated, that the individual himself is the only proper judge of his well-being; that the value judgments of the investigator can thus be minimized; and, perhaps most important, that it is not appropriate to measure well-being in old age by the same standards that apply to middle age, namely, standards based upon activity or social involvement (p. 134).

By relying on the individual’s own evaluations, happiness, and satisfaction, in this view, researchers could avoid multiple problems associated with assessments of success or competence.

There were several studies that relied on this “internal frame of reference” (e.g., Lebo, 1953). Many of the early studies combined elements of the two approaches, and asked questions about “happiness” and “feelings of well-being or satisfaction” along with questions designed to assess degrees of social competence (Pollak, 1948, pp. 66-67). Thus, in an influential 1949 book, Ruth Shonle Cavan, Burgess, Robert J. Havighurst, and Herbert Goldhamer aspired “to define and to analyse the nature, patterns, and problems of personal adjustment to ageing” (Cavan, Burgess, Havighurst, & Goldhamer, 1949, p. v). The second author of this book was also the first author of a major contribution to the literature on marital happiness (Burgess & Cottrell, 1939), which underscores the fact that the different traditions explored here were not independent of one another. Cavan et al. asked their 3,000 participants questions about both whether they felt satisfied, and whether, for example, they participated in clubs and organizations (Cavan et al., 1949, Appendix A).

The most common focus in gerontology, however, was on “life satisfaction” (cf. Adams, 1971; Neugarten et al., 1961). Thus, Arnold M. Rose (1955) studied the life satisfaction of married adults whose children had just left the home. Rose asked his subjects: “In general, how satisfied are you with your life?” and offered them a list of five alternative answers: “Very dissatisfied,” “Somewhat dissatisfied,” “About average,” and so on (Rose, 1955, p. 15). Rose found that women (but not men) who married before the age of 20 or after the age of 30 tended to be less satisfied with their lives, that dissatisfied women were less likely to have a paying job and more likely to think they spent an inordinate amount of time doing housework, and that satisfaction was associated with social participation in both men and women (Rose, 1955, pp. 18-19). A more explicit treatment of measures of life satisfaction appeared in Neugarten et al. (1961). They took satisfaction to be closely related to “Zest (vs. apathy); Resolution and fortitude; Congruence between desired and achieved goals; Positive self concept; and Mood tone” (p. 137), but did not specify if these factors were supposed to be components, causes, or mere correlates of satisfaction.

In sum, gerontologists have been interested in wellbeing since the establishment of the discipline. While some authors in this tradition spoke about “psychological well-being” (Neugarten et al., 1961, p. 134), they frequently talked about “well-being” without the modifier (cf. Adams, 1971, p. 64; Neugarten et al., 1961, p. 134). Numerous gerontologists appear to have turned to happiness and satisfaction as proxies for wellbeing because they thought that by doing so they could avoid imposing their own value judgments and standards of wellbeing on their subjects; because they wanted to rely, to the greatest extent possible, on the subjects’ own standards of wellbeing; and because they felt that the subjects themselves were best situated to

determine to what extent they satisfied those standards. The tension between the two approaches to wellbeing measurement in gerontology is interesting because it mirrors the contemporary tension between subjective measures of wellbeing and capability-based measures. Capability-based measures are designed to reflect not mental states like happiness but capabilities and functionings, that is, what a person can be and do in leading a life (Sen, 2008, p. 24).

2.8 The social indicators movement

A great deal of interest in subjective wellbeing during the 1960s and 70s grew out of the social indicators movement. This movement emerged in the late 1960s as a reaction against the widespread use of economic indicators like GDP as measures of “the goodness of life” (Andrews, 1989, p. 401; Campbell, 1976, p. 118). For instance, Angus Campbell (1976) admitted that economic measures were “easy to count” and he suggested that this goes a long way toward accounting for their popularity as measures of wellbeing. Yet, he added: “None of us doubts that economic data have admirable qualities; the question is, How well do they represent the quality of national life? How valid are they as measures of the goodness of life in this country?” (p. 117). Campbell was obviously not enthusiastic about economic measures. In contrast, the social indicator movement proposed a set of indicators that, although perhaps not as “easy to count,” would nevertheless be better measures of quality of life. As Frank M. Andrews put it:

The idea called for two key changes from earlier practices. One was an expansion in the range of phenomena monitored beyond the traditional economic indicators, and an explicit recognition that “life quality,” however it might be designed, involved more than just economic considerations. The second change involved an attempt to focus directly on “output” indicators – i.e., indicators that show how well off people actually are – in addition to the more traditional “input” indicators that reflect budget allocations, procedures and processes that are presumed to enhance well-being (Andrews, 1989, p. 401).

The social indicator movement instead generated a wealth of so-called “objective indicators,” that is, statistics concerning life expectancy, quality of food and water, access to medical care, level of education, quality of housing, and so on (Campbell, 1976, p. 118). The underlying assumption, of course, was that more education, housing, and medical care tended to increase the quality of life. As Campbell put it: “It is reasonably argued that as the level of education rises, the adequacy of medical care improves, the amount of substandard housing is reduced, and the purity of the air and water is increased, the quality of life is therewith enhanced” (p. 118).

Nevertheless, a number of researchers came to think that objective indicators did not go far enough. As Campbell wrote:

[If] we believe, as I assume that most psychologists do, that the quality of life lies in the experience of life, then these are surrogate indicators. They describe the conditions of life that might be assumed to influence life experience, but they do not assess that experience directly (1976, p. 118).

The problem was that indicators concerning education, housing and health care remained “input” indicators, and as such were subject to the same criticism as economic welfare measures. Thus, the very argument that led to the rejection of economic indicators appears to have led to the rejection of objective indicators as well. In Campbell’s words:

If we are primarily concerned with describing the quality of life experience of the population, we will need measures different from those that are used to describe the objective circumstances in which people live. We will have to develop measures that go directly to the experience itself. These *subjective* measures will surely not have the precision of indicators that are expressed in numbers of dollars, units of time, or numbers of square feet, but they will have the great advantage of dealing directly with what it is we want to know, the individual's sense of well-being (p. 118).

Hence, it appears that the logic of their position forced members of the social indicator movement to reject an exclusive reliance not just on economic measures but on the wider set of "objective" measures as well, and to favor what they called more "direct" measures of quality of life.

One hugely influential study in this vein is Princeton psychologist Hadley Cantril's book *The pattern of human concerns* (1965). Cantril was particularly interested in people's aspirations and in the manner in which aspiration levels affected their happiness and satisfaction. His goal was to find a method to estimate the aspirations and satisfactions by reference to the subjects' own, possibly changing, standards (Cantril, 1965, p. 21). For this purpose, he invented what he called the "Self-Anchoring Striving Scale" (p. 22). When using the scale, the subject is asked to judge how good his or her life is on a graphical, ladder-like device, where the top and bottom rungs represent respectively the best and worst possible life he or she could live (p. 23). Cantril and his collaborators administered the test to a total of 23,875 subjects in 13 countries. Under the heading "Who Are the Satisfied?," Cantril discussed what characterized individuals who rated themselves high on the Self-Anchoring Striving Scale. Among other things, Cantril found that education, income, and occupation were strongly correlated with ladder ratings, that "people living in urban centers" rated themselves somewhat higher than "those living on the land," but that men and women rated themselves about the same (p. 258).

Campbell, Philip E. Converse, and Willard L. Rodgers opened up their discussion in *The quality of American life* (1976) by pointing to deficiencies both with traditional economic measures of wellbeing and with those proposed by the social indicator movement. Their book had two objectives. First, the authors wanted to establish "a system of reporting of psychological data which parallels our current repertoire of social indicators" (Campbell et al., 1976, p. 4). Second, they wished to provide "a fuller and truer representation of the state of society to those people who are responsible for social decisions" (p. 5). Campbell et al. (p. 5) drew primarily on the three following sources: Gurin et al. (1960), which they call the "first major study of the quality of life experience"; Bradburn and Caplovitz (1965); and Cantril (1965). On the basis of personal interviews with a random sample of 2,164 Americans (Campbell et al., 1976, p. 511), the authors reported: "Persons who are currently single generally report a good deal less satisfaction with life than do married persons, and the lack of satisfaction shown by women and men who are divorced or separated is quite remarkable indeed" (p. 36). The young tended to report higher happiness than the old, but the old appeared to be more satisfied with life than the young (p. 36). The unemployed and the divorced scored "conspicuously low" (p. 51) and there was a "rather strong and regular relationship between income and a sense of well-being" (Campbell et al., 1976, pp. 51-55).

Frank M. Andrews and Stephen B. Withey's book *Social indicators of well-being: Americans' perceptions of life quality* (1976) was another major contribution. Andrews and Withey placed their research firmly in the social indicator movement. They wrote that "social indicators of the various conditions of human beings, and of the changes that characterize their lives, offer much

that is attractive,” but added that “in these days of growing interdependence and social complexity we need more adequate cues and indicators of the nature, meaning, pace, and course of social change” (p. 1). Though the two said they chose to concentrate on “perceptions of well-being,” they immediately went on to say: “Measurements of individual well-being seem to us, and to many others also, a particularly promising place to begin” (pp. 6-7). This suggests that they in fact did not draw a sharp distinction between wellbeing on the one hand and perceptions of wellbeing on the other. The study drew primarily on “four nationally representative probability samples of the American adult population,” for a total of approximately 5,500 individuals (p. 24).

The social indicators movement, then, played a critical role in the development of subjective measures of wellbeing. Though prominent people like Andrews and Campbell had initially advocated the use of objective measures, the logic of their position pushed them toward more “direct” measures of happiness and satisfaction. The argument that subjective measures are superior because they are more direct has since become part of a standard defense of subjective measures (cf. Angner, 2011). Like at least some of the gerontologists, these researchers thought of measures of happiness and satisfaction as measures of wellbeing, without any modifiers. This is clear from their use of the term “subjective measures of well-being,” e.g. in the title of Campbell (1976), and from the fact that subjective measures were compared with, and found to be superior to, the more established economic welfare measures and the so-called objective indicators. These authors continued the development of subjective measures, and some of them – like Cantril’s Self-Anchoring Striving Scale – remain in use today. Like the epidemiologists we encountered in Section 2.5, authors like Andrews and Campbell aspired to track the wellbeing of large, representative samples over time. Unlike its predecessors, the social indicator movement unashamedly used subjective measures of wellbeing as macro-level indicators of quality of life – that is, as measures of social wellbeing – comparable to the standard economic indicators. Since the subjective measures were thought to be superior to economic or other objective measures, subjective measures of wellbeing emerged as serious substitutes for, e.g., measures of GDP as welfare indicators. This idea remains part of the effort to develop “national well-being accounts,” as described in Section 2.5.

2.9 Summary and discussion

In this section, I have sketched the emergence and development of measures of happiness and satisfaction. Far from being a novel invention, these measures have an uninterrupted history that can be traced back at least to the 1920s and 30s. It was in the context of marital success studies that researchers like Davis, Terman, and others started systematically studying happiness and satisfaction, although some of them were interested in the happiness of the marriage, rather than that of the individuals in it. Educational psychologists like Watson were unambiguously interested in the happiness and satisfaction of individuals, as opposed to marriages; their use of a wide range of remarkably sophisticated happiness measures allowed them to argue on empirical grounds that simple self-report measures were both reliable and valid and could substitute for more extensive batteries of questions. The confidence that attributes of the person could be reliably and validly measured, and that the results could be useful to governments and corporations, was due to the emergence of personality psychology shortly after World War I. Epidemiologists of mental health started using measures of happiness and satisfaction as measures of mental health in large, nationally representative samples. Meanwhile, gerontologists used, above all, measures of satisfaction as measures not just of mental health, or psychological wellbeing, but of wellbeing without the modifier. The

social indicator movement entrenched the idea that measures of happiness and satisfaction could serve as macro-level indicators of quality of life – that is, as measures of social wellbeing – and as serious substitutes for, e.g., measures of GDP as welfare indicators.

It might be objected that the foregoing discussion ignores interesting and important contributions to the literature. Indeed, several early contributions that were in important respects similar to those discussed above, like Adolf Wohlgemuth (1919), J. C. Flügel (1925), and Arthur E. Morgan (1934), have not been included. Obviously, the story told here is incomplete: given how long and rich the history turned out to be, some omission is inevitable. But it is also worth remembering that my aim was not to identify researchers who “anticipated” modern research efforts, but to try to identify the work that was the most influential. Discussion of the economists who brought happiness research to the attention of mainstream economics, above all Bernard M. S. van Praag (1968) and Richard A. Easterlin (1974) has also been omitted. These more recent developments are not discussed because their influence begins as this story ends.

3. Porter on measurement in social and behavioral science

In this section, I will argue that the history of subjective measures of wellbeing from the 1920s or 30s, in all essentials, confirms Porter’s general account of the history of measurement in social and behavioral science. Porter’s account is most clearly articulated in the book *Trust in numbers: The pursuit of objectivity in science and public life* (1995) but has been discussed and developed in a number of forums, including the collection *The age of economic measurement* (Klein & Morgan, 2001) and the special issue of *Studies in History and Philosophy of Science* (Vol. 32, No. 4, 2001), edited by E. P. Hamm and Alan W. Richardson. Consistent with Porter’s account, I will argue, subjective measures of wellbeing emerged in applied, rather than theoretical, branches of social and behavioral science, where they were developed not as a result of physics envy but of a moral impulse to improve society; quantification was intended to make up for perceived deficiencies in unaided human judgment; and radical disagreements about the nature of wellbeing did not impede efforts to measure it – indeed, in time, there was considerably more agreement about how to measure wellbeing than about how to define it.

3.1 Subjective measures appeared in applied branches of psychology

One of Porter’s main theses is that the drive toward quantification was strongest in applied rather than theoretical branches of social science, and that it was the result not of physics envy, but rather of a moral impulse to understand and improve a changing society (Bateman, 2001, p. 57; Levy, 2001, p. 724; Porter, 1995, p. viii). The history of the measurement of happiness and satisfaction confirms Porter’s thesis: measures of happiness and satisfaction emerged in applied branches of psychology, sociology, and medicine, and they appeared not as the result of an effort to replicate the successes of quantification in the natural sciences, but rather of an ambition to understand the causal antecedents of happiness and satisfaction in order to build a happier and therefore better society.

As we have seen, subjective measures of wellbeing did in fact emerge in applied branches of psychology, sociology, and medicine: in marital success studies (Section 2.1), educational psychology (Section 2.2) and personality psychology (Section 2.3), as well as in the epidemiology of mental health (Section 2.5), gerontology (Section 2.7), and social indicators research (Section 2.8). It is true that happiness, satisfaction, and other positive or desirable emotional states have begun to attract attention from more theoretical branches of psychology, economics, and neuroscience. Hence, a collection like *Well-being: The foundations of hedonic*

psychology (Kahneman, Diener, & Schwarz, 1999) contains chapters, e.g., on the neuroscientific basis for utility appraisals. Yet, this is a relatively recent phenomenon, which does not change the fact that the measures originally emerged in applied branches of social and behavioral science.

There is little evidence that proponents of subjective measures suffered from “physics envy,” in the sense of an urge to replicate the successes of quantification in physics and other natural sciences. Some of Hart’s examples of successful measurement – including the electrocardiograph and the thermometer – did arguably come from the natural sciences; yet, his focus was unambiguously on practical applications in the interest of investigating and solving social problems. Here is how he introduced his discussion about the power of quantification:

When intelligence tests were invented a new instrument was placed in the hands of workers with human beings – a powerful instrument for diagnosing maladjustments, assaying human capabilities, placing students and workers in the surroundings where they would be most useful, and investigating various social problems. When the fever thermometer was invented a new instrument was placed in the hands of physicians – a powerful instrument for diagnosing bodily ailments. When the electrocardiograph was invented a new instrument for tracing the pulsations of the heart was made available for diagnosis and for pointing the way toward remedial treatment (Hart, 1940, p. v).

Hart also pointed to the contribution of “aerial dynamics” to the velocity of human transportation (p. 18); yet, the example was only mentioned in passing and the emphasis on transportation suggests a concern with engineering rather than theoretical physics. Moreover, his most powerful example came from epidemiology. Hart reminded us that reliable statistics about infant death rates were critical in establishing that many infant deaths were the result of contaminated water rather than “summer complaint” or “the will of Heaven,” and that the knowledge derived from such statistics allowed the mortality rate to be dramatically reduced (p. 17). In brief, though Hart mentioned tools like thermometers and disciplines like aerodynamics, he was primarily interested in replicating the successes of applied sciences like epidemiology than of theoretical branches of physics and other natural sciences.

The main impetus behind measures of happiness and satisfaction appears to have been a moral impulse to understand and improve society, specifically, an ambition to identify the causal antecedents of happiness in order to build a happier, and therefore better, society. The ambition is plainly evident throughout the history of measures of happiness and satisfaction. In early literature, authors like Davis, Hamilton, and Terman were clearly trying to identify the factors that made for a happy marriage because they thought that a grasp of these factors would allow them to promote success in marriage. Terman said that he was not committed to the view that “personal happiness is the only ‘proper’ goal of marriage,” but then added: “It is of the very nature of happiness that, other things being equal, it should be preferred to its opposite” (1938, p. 2). And Burgess and Cottrell, as we saw above, believed that marital adjustment and incompatibility had become a social problem and therefore of public concern (1939, p. 1).

Educational psychologists like Watson and Hartmann shared the ambition. As we saw in Section 2.2, Watson said he studied happiness because no other human quest had a larger following, because even “the highest ethics” emphasized the importance of human happiness, and because educational programs were justified only in terms of their contribution to happiness (1930, p. 79). Unlike some of the later authors, Watson did not frame his project in terms of wellbeing or welfare. Yet, his use of the term “highest ethics” suggests that his desire

to explore scientifically the determinants and distribution of happiness may have been due to philosophical or religious commitments to happiness as a central constituent of the good life.

Hartmann (1934) started out by quoting Raymond Dodge, past president of the American Psychological Association. In a brief autobiographical narrative published in 1930, Dodge had written:

I am desirous of participating in the solution of ... the problem of protracted human happiness. Whether or not it fits in with one's philosophy of life, the fact is incontestable that happiness is an important if not the most important aim of human endeavor (Dodge, 1930, p. 119; quoted in Hartmann, 1934, p. 202).³

Hartmann obviously agreed: "If happiness be one of the major goals of living, if not the only consciously acceptable end of life itself," he wrote, "surely an analysis of the conditions fostering or hindering its attainment is an intellectual obligation of the first order" (1934, p. 203). Evidently, both Watson and Hartmann assumed that a happier world would be a better world and wanted to understand happiness in order to promote it.

Hartmann (1934, p. 203) made favorable mention of Abraham Myerson's program of *eupathics* (Myerson, 1917). Myerson was Clinical Director at the Taunton State Hospital in Massachusetts and a Harvard neuropathologist. His goal was to establish a field of "Mental Hygiene" that had "for its aim the *well being of the normal*" (Myerson, 1917, p. 344). He called the program the "more gracious sister" of eugenics, which "largely simmers down to a program for the elimination of the unfit" (p. 344). Myerson had great hopes for the field. He appears to have equated wellbeing with happiness, and believed that "Mood" was an important determinant of happiness. Furthermore, he claimed that Mood "can be reached and elevated in a perfectly definite manner" (p. 344). Myerson invested eupathics with a great deal of importance. After asking rhetorically what was left out of the field, he answered: "nothing of consequence" (p. 346).

Hart and Goldings were equally explicit about their desire to better understand happiness in order to promote it. As we saw in Section 2.4, Hart believed that "[one] of the basic purposes of mankind is to be happy" (1940, p. 16). His Euphorimeter was presented as designed "to point the way, not to administration of drugs, but to constructive measures which may relieve maladjustments, promote the cure of mental suffering, and open the way toward more joyous living" (p. v). The aim was to help people "live joyously within a menacing world ... in spite of threats and pressures of war, of economic disaster, of our own incurable physical handicaps and past emotional wounds, and of the original natures of the people with whom we have to live and work" (p. 6). Similarly, Goldings explained his interest in happiness as follows:

"How to gain, how to keep, how to recover happiness is, in fact, for all men at all times the secret motive of all they do, and of all they are willing to endure." This pithy statement by William James reflects the almost unanimous judgment of prescientific psychologists from Aristotle down. Most laymen, caviling aside, would probably agree with it (Goldings, 1954, p. 30).

Both Hart and Goldings appear to have operated on the assumption that the scientific study of happiness would help us build a happier, and therefore better, world.

³ In passing, Dodge's writings suggest that he found inspiration less from natural science than from philosophy. In the same autobiographical remarks, he wrote: "I am inclined to believe that the study of the history of philosophy and the logic of science form a valuable background for estimating the possibilities and the limitations of experimental evidence" (Dodge, 1930, pp. 120-103).

The epidemiology of mental health is premised on the assumption that a better understanding of the causes and distribution of mental health might allow us to promote it more effectively. Gurin et al. conclude their book by writing: "We hope that this study ... contributes to the formulation of programs designed to make [psychiatric] help available to the maximum number of people" (Gurin et al., 1960, p. 406). Similarly, as we saw in Section 2.7, gerontology had always been centrally concerned with measuring and promoting the wellbeing of older people. Neugarten et al. (1961), for example, were dissatisfied with existing measures based on various criteria of success or competence, which they felt relied on dubious, value-laden assumptions; by developing better measures of wellbeing, they thought that they would be in a better position to understand and promote the wellbeing of older people.

The social indicator movement, as we saw in Section 2.8, was motivated by a desire to develop accurate measures of "the quality of national life," "the goodness of life," and "life quality," that is, "indicators that show how well off people actually are" (Andrews, 1989, p. 401; Campbell, 1976, p. 117). Authors like Campbell et al. charged that widely used economic measures and social indicators were imperfect proxies for wellbeing, and maintained that "the nation must change from its fixation on goals which are basically economic to goals which are essentially psychological, from a concentration on being well-off to a concern with a sense of well-being" (1976, p. 1). They took the importance of happiness and satisfaction to be uncontroversial. As Andrews and Withey put it: "[It] would seem fair to say that the importance of human satisfaction does not need to be defended" (1976, p. 10). The view that a society with more quality of life, wellbeing, etc., would be a better society infused all this work. Hence:

The promotion of individual well-being is a central goal of virtually all modern societies, and of many units within them. While there are real and important differences of opinion – both within societies and between them – about how individual well-being is to be maximized, there is nearly universal agreement that the goal itself is a worthy one and is to be actively pursued (Andrews & Withey, 1976, pp. 6-7).

From the outset, these authors very much wanted their work to inform public policy, as when Campbell et al. said they wished to provide "a fuller and truer representation of the state of society to those people who are responsible for social decisions" (1976, p. 5). Cantril made it clear that his research was ultimately aimed not only at advancing scientific understanding, but also at designing new institutions: "More reliable predictions of what people want or do not want, believe or do not believe, will accept or will not accept, should aid the process of creating new forms of economic, social, and political institutions" (1965, p. 3).

The moral impulse to build a happier and therefore better society is no less evident in contemporary writings on subjective measures of wellbeing. As part of their discussion of national wellbeing accounts, Ed Diener and Seligman write:

The most important contribution of a national system of well-being indicators would be that they could focus the attention of policymakers and the public specifically on well-being, and not simply on the production of goods and services; one of the main benefits of well-being measures is that they add a valuable perspective beyond a cost-benefit market analysis in evaluating societal structures and interventions (2004, p. 21).

The hope is that national wellbeing accounts will help policymakers understand and promote that which really matters, which in their view is subjective wellbeing. The underlying

assumption is that a happier society would be a better society. Similarly, Richard Layard writes that “[Bentham] proposed that all laws and all actions should aim at producing the greatest possible happiness... I believe that Bentham’s idea was right and that we should fearlessly adopt it and apply it to our lives” (2005, pp. 111-112). The central importance of happiness for a good society is thought to entail that public policy should be designed to promote it. Thus, Diener and Seligman “propose that [subjective] well-being ought to be the ultimate goal around which economic, health, and social policies are built” (2004, pp. 1-2). In these passages, by the way, there is no trace of physics envy.

In sum, the history of the measurement of happiness and satisfaction confirms the thesis that measures of happiness and satisfaction emerged in applied branches of psychology, sociology, and medicine. Though happiness and the like have begun to attract attention from more theoretical branches of psychology, economics, and neuroscience, measures of happiness did not emerge there. The history also confirms that measures of happiness and satisfaction did not appear as a result of physics envy; while Hart mentioned tools like thermometers and electrocardiographs, as well as disciplines like aerodynamics, he was interested primarily in practical applications, and he was most impressed by the successes of quantification in applied disciplines like epidemiology. The drive toward quantification was due to a desire to understand the causal antecedents of happiness in order to build a happier and therefore better society.

3.2 Subjective measures as a means to overcome deficiencies in human judgment

Porter has also argued that quantification was intended to make up for perceived deficiencies in unaided human judgment, in particular, a lack of trust and suspicions of arbitrariness (Levy, 2001, p. 724; Porter, 1995, p. 199). This contention helps explain why the drive toward quantification was strongest in ‘soft’ disciplines like psychology in general and their applied subdisciplines in particular. In this view, practitioners in soft disciplines were most insecure, and therefore most susceptible to outside and inside pressure; they turned to empirical research to establish answers that could not be accused of being biased by personal, religious, and other considerations (Porter, 1995, p. 200). The history of measurement of happiness and satisfaction confirms Porter’s contention: throughout the history reviewed here, scientific measurement was seen as a means to overcome the limitations associated with unaided human judgment, in particular, a lack of trust and suspicions of arbitrariness.

Since the very early days of happiness measurement, contributors commonly complained that science had failed to give adequate attention to the determinants and distribution of happiness and satisfaction. Watson’s remark that it was “extraordinary almost beyond belief that so few attempts have been made to apply the techniques of psychological study to the understanding of happiness” (Watson, 1930, p. 79) has echoes throughout this history. For example, the complaint appears in Dodge’s autobiographical sketch. Notwithstanding the importance of happiness, Dodge wrote, “it has received no commensurate scientific attention [sic]. The theory of the happy life remains at about the level where the Greek philosophers left it” (Dodge, 1930, p. 119; cited in Wilson, 1967, p. 302). He continued:

Scientific [data] as to the fundamental positive conditions of protracted happiness are conspicuous for their absence... The positive conditions of happiness are left largely to accident, such as the satisfaction of instinctive wants with its tragic disillusionments and negative adaptations, the economic pressure to provide a market for manufactured products, the exigencies of the labor market, the desire to amass wealth, or the Bolshevistic abolition of private

wealth. There are numerous wise protests that protracted happiness is achieved by none of these things; but positive, scientific data on the real conditions are as inconspicuous as scientific interest in the problem (Dodge, 1930, p. 120).

Dodge believed that a scientific approach based on scientific measurement was more promising than alternative approaches; that in the absence of scientific data, “wise protests” were distrusted and largely impotent; and that using unaided human judgment rather than “the techniques of psychological study” was akin to leaving it to accident. The fear of arbitrariness is highly evident.

The same concern appears in Hart and Goldings’s work. Hart noted:

By measuring accurately, scientists have trebled the length of human life – but then we discover that they have left a lot of people wondering whether life is worth living... What our scientists have been doing has been to use measurement to increase things which people *thought* would contribute to happiness, but without ever measuring happiness itself to see whether the alleged means were actually contributing to the desired end (1940, pp. 18-19).

Quite obviously, Hart did not trust unaided human judgment to reveal what “actually” contributes to happiness. As we saw in Section 2.4, Hart favored a scientific approach. He asked: “Can recent advances in scientific thinking tell us more and more effectively what to do in order to be happy, and in order to help make our fellow human beings happy?” (1940, p. 16). He answered:

The present book is based on two propositions: first, that it is possible to measure happiness and unhappiness reliably; and second, that, if we can thus measure, we can then move on toward discovering the causal factors by means of which we can learn with more and more effectiveness to eliminate misery and increase joy scientifically (p. 16).

Hart believed that a scientific approach to identifying the causes of happiness, and a scientific approach only, would permit happiness scholars to be as effective in enhancing happiness as epidemiology had been in eliminating disease and aerodynamics in building airplanes. Having pointed to the enormous success of measurement in fields like epidemiology, he added: “Why not begin to apply these methods of precise measurement when we come down to the crucial question of happiness?” (p. 18). Goldings (1954) concurred. In spite of the virtual consensus on the importance of happiness, he wrote, “there has been a marked reluctance on the part of present-day researchers to undertake systematic studies of happiness, unhappiness, and related phenomena” (p. 30). When he proceeded to talk about the questions that have to be addressed by “a psychology of happiness,” he clearly assumed that such a psychology would uncover insights that would be unavailable in the absence of scientific study.

As we saw in Section 2.8, authors like Andrews and Campbell were highly critical of economic measures and social indicators because of our ignorance of how objective conditions translated into subjective experience, which, they thought, was what ultimately mattered. As Campbell et al. wrote, “the fact is that we do not know how well objective measures like these represent underlying psychological states or how well social indicators can be taken to represent the quality of life experience” (1976, p. 3). Their research aimed to make up for this deficiency:

The research with which this book is concerned derives from the conviction that the relationship between objective conditions and psychological states is very imperfect and that in order to know the quality of life experience it will be

necessary to go directly to the individual himself for his description of how life feels to him. This obviously will take us into the subjective world of perceptions, expectations, feelings, and values and will involve us in excruciating problems of definition and measurement (p. 4).

While Campbell and his coauthors recognized that a scientific study of subjective phenomena would face difficult problems, they clearly believed that a non-scientific approach would be even less promising.

The theme is echoed in contemporary writings as well. Seligman maintains that when it comes to what makes us happy, “the field of psychology [is] a puzzling disappointment” (2002, p. ix). And Diener and Seligman target public policy when they write:

While wealth has trebled over the past 50 years, for example, well-being has been flat, mental illness has increased at an even more rapid rate, and data, not just nostalgic reminiscences, indicate that the social fabric is more frayed than it was in leaner times. These inadequacies lead us to advocate that an ongoing system of indicators be instituted by governments and organizations to track well-being over time. It is clear that policymakers now care about well-being, in addition to economics, because policies are being created on the basis of mere guesses and romantic sentiments about what will enhance well-being (e.g., family leave). These guesses are undoubtedly correct in some cases, but they are incorrect in other cases. For this reason, ongoing measurement of well-being in representative samples and in diverse domains of life is required to confirm or disconfirm the efficacy of policies instituted to increase well-being (2004, pp. 20-21).

In Diener and Seligman’s view, the problem can be solved by shifting from unaided human judgment to a system of national wellbeing indicators, which would provide the hard data required to eliminate the reliance on unscientific beliefs. Interestingly, proponents of subjective measures have also turned the tables on their critics, who do not trust the validity and reliability of subjective measures, and attacked them for failing to marshal any empirical evidence that undercuts the validity of such measures (Diener, Lucas, Schimmack, & Helliwell, 2009, p. 75). This move, triggered by the charge of arbitrariness and the lack of trust, implicitly faults the critics of subjective measures with relying too heavily on unaided human judgment.

In brief, the history of happiness and satisfaction measurement confirms the contention that scientific measurement was seen as a means to overcome the limitations associated with unaided human judgment, in particular, a lack of trust and suspicions of arbitrariness. Operating on the assumption that a happier society would be a better society and that happiness therefore is worth pursuing (cf. Section 3.1), these researchers doubted that unaided human judgment was sufficient to identify the antecedents of happiness and considered a scientific approach more likely to accomplish this goal. The concern was that hypotheses generated by non-scientists were no better than mere guesses and romantic sentiments and that acting on such hypotheses was tantamount to leaving it to chance. The contention that subjective measures were seen as a means to overcome the limitations associated with unaided human judgment is consistent with the fact that the drive toward measuring happiness and the like appeared in applied subdisciplines like educational psychology, which are most frequently criticized for being ‘soft’ and therefore most vulnerable to a lack of trust and suspicions of arbitrariness.

3.3 Disagreements about the nature of wellbeing

Porter has also proposed that the drive toward measurement is not typically impeded by disagreements about the nature of the object under study (M. S. Morgan, 2001, p. 248; Porter, 1995, pp. 94-95). The point was made in the context of attitude measurement as early as 1935, when Allport noted that “attitudes today are measured more successfully than they are defined” (Allport, 1935, p. 828). The history of happiness and satisfaction measurement confirms Porter’s proposition: the researchers discussed here (insofar as they expressed a view at all) had radically different ideas about the nature of wellbeing, happiness, and satisfaction. Yet, the disagreement was not treated as an obstacle to measurement.⁴

Students of marital success, as we saw in Section 2.1, were deeply interested in happiness, satisfaction, and the like, though their focus frequently was on the happiness of the marriage as opposed to that of the people in it. These figures had little to say about the nature of happiness, satisfaction, and success, though the terms appear to be treated as interchangeable, e.g., when Davis asked whether people considered their lives “happy, satisfactory, successful” or “unhappy, unsatisfactory, unsuccessful” (1929, p. 89). Educational psychologists were no less interested in happiness and satisfaction, though their focus was squarely on individuals rather than on marriages. They too appear to have treated happiness and satisfaction as interchangeable; when Watson (1930) talked about satisfaction as well as happiness – e.g., in the first sentence of his paper – there is no evidence that he used the terms in any other way than as synonyms. He did say: “What is studied might, in the strictest sense be termed not happiness, but self-estimates of happiness” (p. 79). Yet, the fact that from that point on he talked about happiness suggests that in Watson’s mind the two were at least tightly linked.

Terman (1938) eschewed definitions. In order to avoid “philosophical connotations,” he wrote, he “preferred to apply the term in a sense familiar to everyone” (p. 3) without specifying what that sense was. Terman immediately went on to offer what appears to be a purely operational definition: “The important thing to remember is that throughout this discussion happiness or unhappiness will mean a high or low score on a particular scale” (p. 4). The scale in question was constructed by combining subjective ratings with other “more objective” judgments, but from the weighting scheme it is clear that he considered the subjective rating the most important item. Sometimes, Terman (e.g. on p. 367) referred to his marital happiness score as an index of satisfaction, which suggests that he treated “happiness” and “satisfaction” as interchangeable.

Burgess and Cottrell relied on *Webster’s New International Dictionary*, which defined “happiness” as “a state of well-being characterized by relative permanence, by dominantly agreeable emotion ranging in value from mere contentment to positive felicity, and by a natural desire for its continuation” (Burgess & Cottrell, 1939, p. 31). As the authors note, the dictionary definition emphasizes the subjective nature of the concept. Interestingly, they made no attempt at explaining the meaning of the word to their subjects, on the assumption that people in general would understand the notion in accordance with the dictionary definition (p. 31). Like Terman, Burgess and Cottrell also sometimes talked about “satisfaction” and “success” instead of “happiness” (p. 45).

Hart (1940) appears to have been somewhat ambivalent, as he offered two different definitions of “happiness.” According to the first: “Happiness is the state in which people are when they say sincerely, ‘I am happy,’ and it is the opposite of the state in which they are when

⁴ See Angner (2010) for a more careful review of accounts of well-being implicit in the literature on subjective measures.

they say sincerely, 'I am unhappy'" (p. 182). He added: "It has been assumed that the average person is sincere in his answers to such a test as the Euphorimeter, especially when he does not sign his name ... Moreover, it has seemed reasonable to assume that whatever insincerities do occur tend to cancel each other out" (p. 182). He noted that the first definition, under the assumption of sincerity, "serves operationally as a basis for measurement, but it is rather deficient in providing insight," and "does not take us very far into the meaning of the term" (p. 183). Thus, he proposed a second definition: "Happiness is any state of consciousness which the person tested seeks to attain or to maintain, and it is the opposite of any state which the possessor seeks to change or from which he seeks to escape or withdraw" (p. 183). In the end, however, Hart argued that "the two definitions really define essentially the same thing" (p. 184). The name of the Euphorimeter suggests that Hart also identified happiness and euphoria.

Like Hart, Goldings was unusually explicit about "the intrinsic nature and characteristics" of happiness. He wrote: "Happiness and unhappiness may be considered as zones on a continuum of hedonic affect which embraces feelings of elation, contentment, satisfaction, and pleasure at the positive pole and feelings of depression, discontent, and unpleasure at the negative pole" (Goldings, 1954, p. 31). He added:

[Although] in common parlance, "happiness" refers only to the positive elements of hedonic feeling, in the present investigations "happiness" more generally refers to the entire continuum as a whole and "satisfaction" and "dissatisfaction" refer more specifically to the polar regions of positive and negative hedonic feeling (p. 31).

From this quote, it appears that Goldings drew a sharp distinction between "happiness" as the continuum of hedonic feeling and "satisfaction" as the positive end of the continuum. It is unclear whether he in fact used the terms in this way, however. Consider:

In order to bring happiness and unhappiness (or satisfaction and dissatisfaction) into proper perspective *vis-a-vis* other general areas of psychological study, they may be considered hypothetical physiological states with both *subjective manifestations* (diffuse and pervasive feelings of pleasantness and satisfaction, and of unpleasantness and dissatisfaction) and *objective signs* (posture, gait, facial expression, tone of voice; feelingful verbal statements; indirect or projective indices of feeling) (p. 31).

Here, happiness and satisfaction were both treated as physiological states that cause subjective feelings as well as outward behaviors of a certain kind.

Campbell and colleagues (1976) differed from their predecessors by drawing a sharp distinction between "happiness" and "satisfaction." Having asked questions about satisfaction alongside questions of happiness, Campbell et al. realized that answers to questions of satisfaction were not as highly correlated with answers to questions about happiness as they would have thought (p. 8). They wrote: "This means that there is some significant minority of persons who report relative happiness along with relative dissatisfaction with their lives, and the converse" (p. 9). As a result, they concluded that happiness and satisfaction were two different traits, both of which were constitutive of wellbeing, and which had to be measured independently. "Satisfaction implies a judgmental or cognitive experience," they maintained "while happiness suggests an experience of feeling or affect" (p. 8). The weak correlation between happiness and satisfaction, they continued, also means that the choice of measure may affect the conclusions of a study. Campbell et al. favored a measure of satisfaction – as opposed to happiness – as a measure of the quality of life experience. They offered several reasons. First,

they took “satisfaction” to be easier to define and to translate than “happiness” (pp. 7-8). Second, they wanted their data to be relevant to public policy, and they thought legislators and decision-makers were more accustomed to thinking in terms of satisfying needs than of promoting happiness (pp. 8-9). Third, Campbell et al. wanted a concept appropriate to use both in the context of “life as a whole” as well as in the context of more limited domains (p. 9).

Bradburn and Caplovitz (1965) made another unexpected discovery, viz., that positive affect was remarkably weakly correlated with negative affect. In order to develop what they called “more detailed measures of well-being,” and using a strategy reminiscent of Watson’s adjective-based measure, Bradburn and Caplovitz offered twelve different descriptions, half of them positive, half negative, and asked subjects how often they had felt that way during the past week (pp. 15-16). They wrote: “We expected that the items would cluster in two groups, one indicative of positive and the other of negative feelings, and that the two clusters would be negatively related to one another” (p. 16). As expected, the authors did find “a strong tendency for most of the items to fall into two clusters of positive and negative feelings” (p. 16). However, they were surprised to learn that “items in one cluster are not negatively related to those in the other cluster in any consistent or strong fashion” (p. 18). The two items which were the most strongly (negatively) correlated were “depressed or very unhappy” and “on top of the world,” and these items had a correlation of a mere $-.19$ (p. 18). Interestingly, while positive and negative items did not correlate with each other, each was correlated with happiness (pp. 18-19). The authors inferred that “happiness is a result of the relative strengths of positive and negative feelings, rather than of the absolute amount of one or the other” (p. 21). In light of this finding, it is sometimes suggested that wellbeing should be seen as having three irreducible components: the presence of positive affect, the absence of negative affect, and satisfaction.

In sum, the history of the measurement of happiness and satisfaction confirms Porter’s proposition that the drive toward measurement is not typically impeded by disagreements about the nature of the object under study: the researchers discussed here (insofar as they expressed a view at all) had radically different ideas about the nature of happiness and satisfaction, but the disagreement was not treated as an obstacle to measurement. Many researchers in this tradition declined to discuss what they meant by “happiness” and “satisfaction.” Some of them simply noted that they use the concept in the “established” sense, and that they assumed their subjects to do so too; some made it explicit that they hoped to avoid philosophical discussion. When authors did offer explicit definitions of “happiness,” “satisfaction,” and “well-being,” not only did they disagree with each other regarding the proper understanding of these terms, but there are passages in which some (like Hart and Goldings) appear unable to agree with themselves. Moreover, the perceived relationship between happiness and satisfaction shifted during the period under study. While the early contributors treated happiness and satisfaction as largely interchangeable, Campbell et al. thought of them as different and of wellbeing as constituted by both happiness and satisfaction, and Bradburn and Caplovitz’s results have inspired the view that wellbeing consists of three irreducible components: the presence of positive affect, the absence of negative affect, and satisfaction. Yet, the highly evident disagreement about the nature of happiness was not treated as an obstacle to measurement.

3.4 Convergence in measurement

In light of the disagreement about the nature of the object under study, it is interesting to note that, at least during the episode surveyed in this paper, there was a certain degree of convergence regarding how to measure happiness and satisfaction. It is true that since the

1920s and 30s a variety of instruments have been used to gauge people's happiness and satisfaction; at no point was there consensus about the proper way to measure wellbeing. Yet, there was a clear tendency to substitute a direct question like that of Gurin et al. (1960) for more complex questionnaires. In addition, some found it more convenient to use preexisting measures than to develop their own. Consequently, in time, there was considerably more agreement about how to measure wellbeing than about how to define it.

Watson (1930), as pointed out in Section 2.2, used a surprising variety of remarkably sophisticated measures to assess his subjects' happiness. First, he asked: "Comparing yourself with other persons of the same age and sex how do you feel you should rate your own general happiness?" (p. 79). The subjects were then given a horizontal line, with captions that read, from left to right, "Most miserable of all," "About three-fourths of the population happier than you are," "The average person of your own age and sex," and so on (p. 80). Subjects were asked to make a short vertical mark at the point where they thought they belonged, taking into account their "average state over several months" (p. 80). The subjects were also asked to place a circle at the point where they thought their friends would rate them (p. 80). Second, subjects were offered ten descriptions, of one or two sentences each, and asked: "Among the following descriptions, arranged in miscellaneous order, choose the one which comes nearest to fitting you" (p. 80). The descriptions included the following: "Cheerful, gay spirits most of the time. Occasionally bothered by something but can usually laugh it off," "Ups and downs, now happy about things, now depressed. About balanced in the long run," and "Life often seems so worthless that there is little to keep one going. Nothing matters very much, there has been so much of hurt that laughter would be empty mockery" (p. 81). Third, subjects were given a blank space and asked: "Now write in a sentence or two, something like those above, which you believe will most truly describe your own general happiness in life" (p. 81). Fourth, subjects were given a list of fifty properties – "Enthusiastic," "Troubled," "Annoyed," and so on – and asked: "Check every term which you believe could fairly be applied to yourself in your prevalent attitudes" (p. 81). Half of the descriptions were positive, half negative (p. 81). Finally, using a graphic rating device as in the first question, subjects were asked about their happiness in different areas and in different stages of life (p. 82).

On the basis of subjects' responses, Watson computed an aggregate "happiness score" (p. 82). To check for reliability, Watson used the responses to different items to create two different happiness scores. He noted that the correlation between the scores for men and women were as high as .83 and .85 respectively. As a result of these calculations, he concluded: "This indicates that the measure is reasonably consistent, throughout" (p. 83). Watson also noted that the correlation coefficient between the subjects' own rating of their happiness on a graphical scale and the aggregate happiness scores for men and women were .81 and .82 respectively (p. 86). Thus, he inferred that for some purposes, self-ratings may suffice. He wrote: "For the purpose of separating high and low groups such a simple indicator would probably be adequate" (p. 86). Based on his own results, Sailer drew the same conclusion: "In general, self-estimate of happiness proved to be a characteristic easily measured and showing considerable consistency and relationship to many other sorts of factors" (1931, p. 100).

Hart (1940) also used a range of measures. For his "At-the-Moment Euphorimeter," Hart modified one of the kinds of test used by Watson (1930) and Sailer (1931). Participants were presented with a list of adjectives and asked to underline each adjective "which fairly well describes the way you have felt a good deal of the time or several times since you woke up this morning" and to cross out each adjective "which does NOT describe the way you have felt at any time today" (Hart, 1940, p. 114). The adjectives were chosen by picking all synonyms and

antonyms of “happy” and “unhappy” in a dictionary, then adding all synonyms and antonyms of those, and so on, until he had a list of 48 adjectives (pp. 176-177). A subject was said to have a happy reaction each time he or she underlined a positive adjective or crossed out a negative one, and an unhappy reaction each time he or she underlined a negative adjective or crossed out a positive one (pp. 115-116). The test score could then be computed based on the number of happy and unhappy reactions (p. 177). After administering the test to 2,200 subjects, Hart assured himself that the test was reliable by computing scores separately for each half of the adjectives and finding the reliability of the entire test to be .90 (pp. 178-179). For the “Long-Run Euphorimeter” Hart borrowed a number of questions from earlier studies and added some “which seemed likely to be valuable” (p. 181). Many of the questions asked subjects how they “usually” felt, rather than how they felt at the moment (p. 111). The scoring system for the Long-Run Euphorimeter was calibrated by comparing the results of 2,200 subjects who took both the Long-Run and the At-the-Moment test (p. 181).

The validity of his measures was assured, Hart believed, by the manner in which he had defined “happiness.” In effect, he argued, “the adjective battery consists in offering an opportunity to say ‘I am happy’ in 48 different ways, or to say ‘I am unhappy’ in 48 different ways, or to make whatever compromise between these extremes fit one’s own state at the time tested” (p. 182). The Long-Run test also contained questions concerning how subjects felt about the change or maintenance of current conditions. Hart wrote:

At every crucial point in a long series of statistical analyses, based on these data, we have found that persons that say, in one form or another, “I am happy,” tend also to say, in various ways, “I want to keep my present way of life developing as it is now going” ... The correlation is so close that it seems wisest to proceed upon the assumption that our two definitions of happiness really define the same thing (p. 184).

He concluded:

Various improvements need to be made, and will continue to be made, in the Euphorimeter tests. But it is believed that, as they now stand, with their present scoring methods, the Euphorimeters measure happiness and unhappiness with sufficient reliability and validity to identify outstandingly happy and unhappy people and to give important aid in the process of increasing the happiness of those tested (pp. 184-185).

Hence, Hart convinced himself on empirical grounds that the two definitions of happiness were interchangeable, and that the two Euphorimeters – both of which were based on various kinds of self-reports – generated measures of happiness that were reliable and valid.

Goldings participated in a study on “personality development and projective techniques” (1954, p. 30), and as we saw, developed some original methods to assess happiness. First, his subjects were shown 30 photographs of faces with ambiguous facial expressions, and asked “to rate the satisfaction or dissatisfaction (happiness or unhappiness) of the people” (p. 34). He added: “The Ss [subjects] rated each picture twice on a 10 point scale (extremely unhappy to extremely happy). The sum of the 60 ratings (2 ratings for each of the 30 pictures) constituted the S’s ascribed-happiness score” (p. 34). The reason why Goldings included this test was the idea that “direct avowal of happiness may tend to be subjected to some distortion,” and that it would be important to have alternative, indirect means of measuring happiness. The significance of the happiness ascribed to others with ambiguous facial expressions is that the subject would be expected, on theoretical grounds, to project his or her own happiness onto the

faces (p. 33). Second, subjects were given linear-type rating scales like those used by Watson (1930) and Sailer (1931) (Goldings, 1954, pp. 34-35). On the basis of the results, Goldings computed an "avowed happiness" score ranging from 1 to 6 (p. 35). Third, Goldings asked five experimenters to rank-order the subjects "on 'general happiness and overall satisfaction' based on the experimenter's subjective, clinical judgment of the individual"; the ratings were retrospective, and based on the impression made by the subjects during previous testing periods (p. 35).

Goldings found that there was "highly significant agreement among the five judges," and that there was "a fairly close general agreement between the S's avowal of his own happiness and the ratings assigned by the clinical judges" (1954, p. 40). In contrast, he found no positive correlation between avowed happiness and projected happiness. A closer examination of the data suggested to Goldings that those subjects whose avowed happiness scores were in the normal range indeed tended to attribute their own mental state to the ambiguous faces, whereas those subjects who were extremely happy or unhappy tended to attribute the opposite state (p. 42). Hence, he dismissed the indirect measure as a measure of happiness, noting that "the general hypothesis of a general supplementary projection of happiness must be rejected" (p. 41).

Bradburn and Caplovitz (1965) agreed with Goldings (1954) that self-report measures were no less adequate than any other kind of measure. In their view, the fact that no other measure had been shown to be decidedly more valid or reliable, entailed that psychological wellbeing may as well be assessed by simple self-report measures. They wrote:

[There] is no evidence that self-reports are any less (or for that matter more) valid than expert ratings or psychological tests for rating people on a mental health dimension. Furthermore, self-reports have the eminently practical virtues of face validity, directness, and ease of use (Bradburn & Caplovitz, 1965, p. 7).

The authors were aware, of course, that there was no agreement on what to call the dimension of interest (p. 1). Yet, in light of all the evidence, they could identify no reason to favor more complex measures over self-reports. In practice, Bradburn and Caplovitz relied heavily on the question from Gurin et al. (1960).

Neugarten et al. constructed two indices of life satisfaction of their own. The first (*Life Satisfaction Index A*) consisted of 20 statements of the form "I've gotten pretty much what I expected out of life" and checkboxes for "AGREE," "DISAGREE," and "UNSURE" (1961, p. 141). The score was computed by awarding the respondent one point every time he or she agreed with a statement indicating satisfaction or disagreed with one indicating dissatisfaction. The second index (*Life Satisfaction Index B*) consisted of twelve questions such as "How satisfied would you say you are with your way of life?" and three options "very satisfied" (for two points), "fairly satisfied" (one point), and "not very satisfied" (zero points) (p. 142).

When confronted with Cantril's "Self-Anchoring Striving Scale", the subject was asked: "When you think about what really matters in your own life, what are your wishes and hopes for the future? In other words, if you imagine your future in the *best* possible light, what would your life look like then, if you are to be happy?," and similarly for the *worst* possible life (1965, p. 23). Thereafter, the subject was shown a non-verbal device referred to as "the ladder of life," a drawing of a ladder-like shape where the rungs had been numbered from 0 to 10 (p. 22). Next, the subject was told: "Here is a picture of a ladder. Suppose we say that the top of the ladder (POINTING) represents the best possible life for you and the bottom (POINTING) represents the worst possible life for you," and asked "Where on the ladder (MOVING FINGER RAPIDLY UP AND DOWN THE LADDER) do you feel you personally stand at the *present* time?" (p.

23).⁵ Cantril's basic idea was that this device would allow him to judge a person's circumstances from the person's own point of view. In this test, Cantril wrote: "A person is asked to define on the basis of *his own* assumptions, perceptions, goals, and values the two extremes or anchoring points of the spectrum on which some scale measurement is desired" (p. 22). To his surprise, Cantril learned that in some cultures ladders are uncommon; thus, he sometimes substituted a picture of a mountain for the ladder (Andrews & Robinson, 1991, p. 73).

Campbell et al. (1976) asked questions about subjects' satisfaction in specific domains – including marriage, job, housing, etc. – as well as about their satisfaction with "life in general" and about their happiness, using the question from Gurin et al. (1960). Finally, Campbell et al. computed what they called an "Index of General Affect" on the basis of "a series of descriptive adjectives" that subjects could use to describe their lives (1976, p. 13). The idea was to explore how domain satisfaction related to general life satisfaction, and how the latter related to other measures of the "quality of life experience." For a great many of the comparisons, however, Campbell et al. relied on a composite "Index of Well-Being," which was "a single variable representing a rather global sense of well-being," and which combined the satisfaction score and the result on the adjective test (p. 49).

Andrews and Withey (1976, Ch. 6) tested 68 different measures but ended up favoring one they thought had more desirable qualities than the others, viz., the *Delighted-Terrible (D-T) Scale*, which consists of seven categories ranging from "Delighted," though "Pleased," and "Mostly satisfied," to "Mixed (about equally satisfied and dissatisfied)" to "Mostly dissatisfied," "Unhappy" and "Terrible" (p. 18). The subjects were given three additional response options: "Neutral (neither satisfied nor dissatisfied)," "I never thought about it," and "Does not apply to me" (pp. 18-19). Andrews and Withey offered a variety of reasons to think that people make the kind of global judgments that their research presupposed and that their measures were valid. One of the reasons they thought it was meaningful to ask for global assessments was that people articulate such judgments "promptly and with apparent ease" (p. 64). Moreover, in their sample, less than one percent checked the box marked "I never thought about it" when asked how they felt about their life as a whole (p. 64). Finally, the authors suggested, the fact that questions such as "How are things?" are so common suggests that people think in those terms (pp. 64-66).

As this review makes clear, since the 1920s and 30s, a variety of instruments have been used to gauge people's happiness and satisfaction; at no point was there consensus about the proper way to measure wellbeing. Nevertheless, at least during the episode surveyed in this paper, there was a certain degree of convergence regarding how to measure happiness and satisfaction. Many of the early contributors, like Watson, Sailer, and Hart used a remarkable variety of measures of happiness and satisfaction. However, over time there was a clear tendency to substitute a direct question like that of Gurin et al. (1960) for more complex questionnaires (cf. Diener et al., 1999, p. 277). Having compared the performance of a variety of measures, Watson, Sailer, and many others concluded that self-reports appeared to perform as well as any other measure and sometimes better; certain measures, such as Golding's projection-based measures, appear to have been conclusively rejected. In addition, some contributors found it useful to employ preexisting measures rather than to develop their own. Among other things, as Wilson (1967, pp. 304-305) had pointed out, the use of the same tool as earlier researchers permits the construction of longer time series data. While some

⁵ The parenthetical notes in small caps are Cantril's instructions to the experimenter (Cantril, 1965, p. 22).

disagreement remained, much of it appears to reflect a disagreement about the nature of wellbeing (cf. Section 3.3) rather than a disagreement about how best to measure it. Either way, in time, there was considerably more agreement about how to measure wellbeing than about how to define it.

3.5 Summary and discussion

In this section, I have argued that the history of subjective measures of wellbeing from the 1920s or 30s, in all essentials, confirms Porter's general account of the history of measurement in social and behavioral science. Consistent with Porter's account, subjective measures of wellbeing emerged in applied branches of psychology, sociology, and medicine, and they appeared not as the result of an effort to replicate the successes of quantification in the natural sciences, but rather of an ambition to understand the causal antecedents of happiness and satisfaction in order to build a happier and therefore better society. Throughout the history reviewed here, scientific measurement was seen as a means to overcome the limitations associated with unaided human judgment, in particular, a lack of trust and suspicions of arbitrariness. The researchers discussed here (insofar as they expressed a view at all) had radically different ideas about the nature of wellbeing, happiness, and satisfaction, but the disagreement was not treated as an obstacle to measurement. Because of a tendency to substitute a direct question for more complex questionnaires, and because some found it useful to employ preexisting measures, in time, there was considerably more agreement about how to measure wellbeing than about how to define it. As a result, some of the self-report measures discussed here remain commonly used. This is true, for instance, in the case of the measures used by Andrews and Withey (1976), Cantril (1965), and Gurin et al. (1960).

4. Conclusion

In this paper I have traced the historical roots of measures of happiness, satisfaction, and the like. The history turned out to be longer than is typically suggested. Far from being a recent development, subjective measures can be traced back at least to the 1920s and 1930s. The appearance of a science of happiness was perhaps marked by Myerson's 1917 study promoting the discipline of eupathics, the "more gracious sister" of eugenics, focusing on "the *well being of the normal*" (Myerson, 1917, p. 344; cf. Section 3.1). Systematic happiness research took off shortly thereafter. Subjective measures emerged in marital success studies, educational psychology, and personality psychology in the 1920s and 30s, as tools developed for the study of personality were applied to happiness and satisfaction; the measures were further shaped by the epidemiology of mental health, gerontology, and the social indicator movement in the 1960s and 70s, as measures of happiness and satisfaction were recruited as proxies for mental health and wellbeing in large, representative samples. The history is not only longer, but also richer, than typically suggested. As we have seen, the researchers in the traditions discussed here had rather different purposes, coming as they did from a variety of backgrounds and being involved in a variety of research projects. Social and behavioral scientists who studied happiness, satisfaction, wellbeing, and the like operated with a variety of definitions. The scientists invented a fascinating variety of measurement tools, some of which were ultimately rejected, but some of which – like the question of Gurin et al. (1960) – turned out to have a remarkable longevity. A number of the studies were longitudinal; some used remarkably large, representative samples.

The history of the measurement of happiness and satisfaction is consistent with Porter's general account of measurement in social and behavioral science. One of Porter's main theses is

that the drive toward quantification emerged in applied rather than theoretical branches of social science, and that it did so not as a result of physics envy, but rather of a moral impulse to understand and improve a changing society. Indeed, measures of happiness and satisfaction emerged in applied branches of psychology, sociology, and medicine, and they appeared not as the result of an effort to replicate the successes of quantification in the natural sciences, but rather of a desire to understand the causal antecedents of happiness in order to build a happier and therefore better society. Porter also argued that quantification often appeared as a means to overcome deficiencies in human judgment, in particular, a lack of trust and suspicions of arbitrariness. In fact, throughout the history reviewed here, scientific measurement was seen as a means to overcome the limitations associated with unaided human judgment, in particular, a lack of trust and suspicions of arbitrariness. In addition, Porter maintained that the drive toward measurement is not typically impeded by disagreements about the nature of the object under study. As it happens, the researchers discussed here (insofar as they expressed a view about the nature of happiness and satisfaction at all) had radically different ideas about the nature of wellbeing, but such disagreement was not treated as an obstacle to measurement – indeed, in time, there was considerably more agreement about how to measure wellbeing than about how to define it.

The story confirms Chang’s thesis that history of science as complementary science can serve multiple purposes (2004, p. 249). First, the story offers further proof that history can serve to recover knowledge about natural and social phenomena as well as about previous efforts to measure them that appears to have been forgotten. The variety and sophistication of the measures used even early in the game, by researchers like Watson and Hart, is astonishing and, it appears, largely forgotten by practicing social and behavioral scientists. Recent work has noted a surprisingly weak correlation between objective conditions and subjective wellbeing (Argyle, 1999). But this phenomenon was well known throughout the history reviewed here; Symonds, for example, concluded: “Happy and unhappy [people] are remarkably alike in their problems and interests. The unhappy do not have peculiar problems but make less satisfactory adjustments to their problems” (1937, p. 293). Meanwhile, recent years have seen a sharp debate about the extent to which people adapt to objective circumstances. Under the heading “Myths in the science of happiness,” Diener criticizes what he calls the “extreme setpoint theory” and argues that circumstances do matter and that adaptation is not in general complete (Diener, 2008, pp. 494-496). But already, in 1940, Hart had shown that recently married couples, prisoners working on roads, and couples who have contemplated divorce exhibit sharply different happiness scores, which would disconfirm the extreme setpoint theory (see Figure 1).

The story also confirms that history can serve to develop critical awareness of contemporary efforts, their prospects and limitations. Among other things, the historical overview offered here sheds light on the methodological and philosophical underpinnings of the project. As one example, critics of subjective measures sometimes suggest that proponents of subjective measures of wellbeing illegitimately assume without argument that people are capable of accurately reporting how happy or satisfied they are. This impression is sometimes encouraged by the proponents themselves. Consider again Frey and Stutzer, who state: “In general, it can be assumed that [individuals] are the best judges of when they are happy and when they are unhappy” (2002, p. 4). They support this assumption by referring to “a sensible tradition in economics to rely on the judgment of the persons directly involved” (p. 4). This move is surprising, given that only a few pages earlier they attack the track record of economics in saying something useful about happiness (p. vii). However, as we have seen, it is a mistake to think that psychologists and economists working on subjective wellbeing simply take

people's ability to report their happiness for granted. Over the course of the history reviewed here, substantial efforts have gone into showing that the relevant measures are reliable and valid (cf. Angner, forthcoming).

Finally, the story demonstrates that history can serve to suggest novel experimental and theoretical developments. As we saw in Section 2.6, Wilson ended his review by calling for more experimental research into the determinants of happiness. While the specific intervention he discussed may not be the most suitable one, he was quite right that the possibility of exploring experimentally the determinants of happiness is exciting. Moreover, Campbell et al. identified interesting avenues of theoretical development when they noted that degrees of satisfaction differ radically between individuals even under objectively identical circumstances. They also pointed out that this fact is part of the *raison d'être* of the whole endeavor: "If there were a close and universal relationship between the level of material possessions and the quality of life experience, there would, of course, be little point in undertaking a study of the kind in which we are here engaged" (1976, p. 10). Campbell et al. suggest that satisfaction may be a function not just of actual or perceived attainment, but also of the standard against which attainment is assessed. Hence:

The individual's assessment may derive from any or all of the following bases of evaluation: *aspiration levels*, or the situation that a person hopes eventually to attain, where a given domain is concerned; *expectation levels*, or the situation he feels he is likely to attain in the fairly immediate future; *equity levels*, or what he thinks should be true of his situation if perfect justice prevails, given how much he invests in it relative to others; *reference group levels*, or what he believes to be true of the situations of others with whom he identifies, such as friends and family or others of his income, race, or occupation; *personal needs*, or the amount of a particular reward he may require, such as how much savings to feel secure, how much housing to be comfortable, how much police protection to feel safe; and *personal values*, concerning such intangibles as freedom, equality, and the like (Campbell et al., 1976, p. 14).

While recent research has explored precisely this kind of question, the area is far from exhausted.

It is my hope that this history, though short and in many ways incomplete, will be of interest not just to historians of psychology but to practicing social and behavioral scientists, as well as to anyone who is interested in the application of the science of happiness to public policy. This historical review has offered new perspectives on past and present scientific practice and further proof that history can serve to recover knowledge – about natural and social phenomena as well as about previous efforts to measure them – that appears to have been forgotten; to develop critical awareness of contemporary efforts, their prospects and limitations; and to suggest novel experimental and theoretical developments. Further study into the history of subjective measures thus seems highly motivated, not just because of what we can learn about the historical origins of such measures but also because of what we can learn about their philosophical and methodological underpinnings.

Acknowledgements

This paper is based on Angner (2005). I am grateful to those, too numerous to name, who have offered feedback on earlier drafts. All errors remain my own. Unless otherwise noted, all italics in original.

Author

Erik Angner
University of Alabama at Birmingham
angner@uab.edu

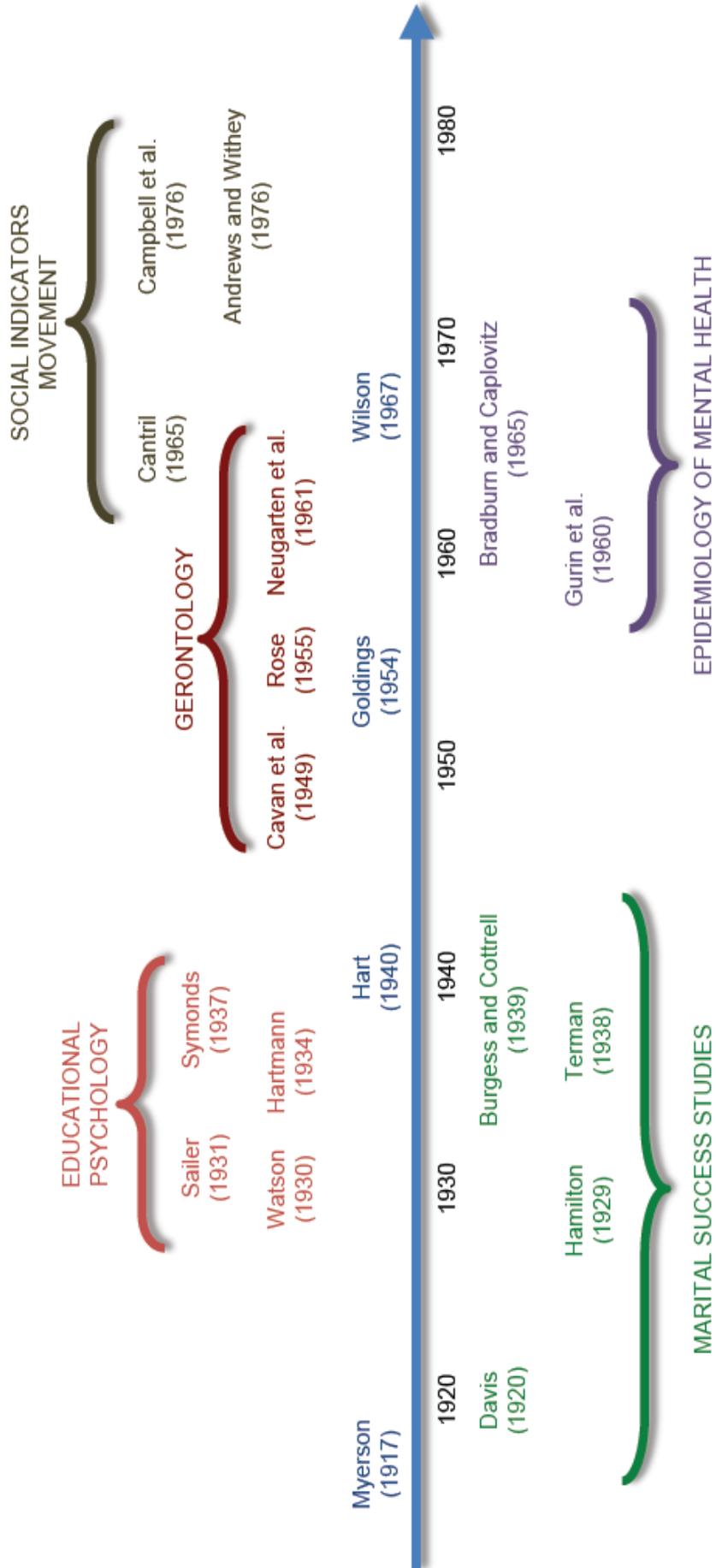
References

- Adams, D. L. (1971). Correlates of satisfaction among the elderly. *The Gerontologist*, 11(4, Pt. 2), 64-68.
- Allport, G. W. (1935). Attitudes. In C. Murchison (Ed.), *A handbook of social psychology* (pp. 798-844). Worcester, MA: Clark University Press.
- Allport, G. W. (1937). *Personality: A psychological interpretation*. New York, NY: H. Holt and Co.
- Andrews, F. M. (1989). The evolution of a movement. *Journal of Public Policy*, 9(4), 401-405. <http://dx.doi.org/10.1017/S0143814X00008242>
- Andrews, F. M., & Robinson, J. P. (1991). Measures of subjective well-being. In J. P. Robinson, P. R. Shaver, L. S. Wrightsman, & F. M. Andrews (Eds.), *Measures of personality and social psychological attitudes*, Measures of Social Psychological Attitudes Series (Vol. 1). San Diego, CA: Academic Press.
- Andrews, F. M., & Withey, S. B. (1976). *Social indicators of well-being: Americans' perceptions of life quality*. New York, NY: Plenum Press.
- Angner, E. (2005). *Subjective measures of well-being: A philosophical examination*. University of Pittsburgh, Pittsburgh, PA. <http://etd.library.pitt.edu/ETD/available/etd-08192005-130324/>
- Angner, E. (2010). Subjective well-being. *Journal of Socio-Economics*, 39(3), 361-368. <http://dx.doi.org/10.1016/j.socec.2009.12.001>
- Angner, E. (2011). Are subjective measures of well-being "direct"? *Australasian Journal of Philosophy*, 89(1), 115-130. <http://dx.doi.org/10.1080/00048400903401665>
- Angner, E. (forthcoming). Current trends in welfare measurement. In J. B. Davis & D. W. Hands (Eds.), *The Elgar companion to recent economic methodology*. Northampton, MA: Edward Elgar. <http://ssrn.com/abstract=1680053>
- Argyle, M. (1999). Causes and correlates of happiness. In D. Kahneman, E. Diener, & N. Schwarz (Eds.), *Well-being: The foundations of hedonic psychology* (pp. 353-373). New York, NY: Russell Sage Foundation.
- Bateman, B. W. (2001). Make a righteous number: Social surveys, the men and religion forward movement, and quantification in American economics. In J. L. Klein & M. S. Morgan (Eds.), *The age of economic measurement*, Annual Supplement to Volume 33, History of Political Economy (pp. 57-85). Durham, NC: Duke University Press.
- Bradburn, N. M., & Caplovitz, D. (1965). *Reports on happiness: A pilot study of behavior related to mental health*. Chicago, IL: Aldine.
- Burgess, E. W., & Cottrell, L. S. (1939). *Predicting success or failure in marriage*. New York, NY: Prentice-Hall.
- Campbell, A. (1976). Subjective measures of well-being. *American Psychologist*, 31(2), 117-124. <http://dx.doi.org/10.1037/0003-066X.31.2.117>
- Campbell, A., Converse, P. E., & Rodgers, W. L. (1976). *The quality of american life: Perceptions, evaluations, and satisfactions*. New York, NY: Russell Sage Foundation.
- Cantril, H. (1965). *The pattern of human concerns*. New Brunswick, NJ: Rutgers University Press.
- Cavan, R. S., Burgess, E. W., Havighurst, R. J., & Goldhamer, H. (1949). *Personal adjustment in old age*. Chicago, IL: Science Research Associates.
- Chang, H. (2004). *Inventing temperature: Measurement and scientific progress*. Oxford: Oxford University Press. <http://dx.doi.org/10.1093/0195171276.001.0001>
- Davis, K. B. (1929). *Factors in the sex life of twenty-two hundred women*. New York, NY: Harper & Brothers.
- Diener, E. (2006). Guidelines for national indicators of subjective well-being and ill-being. *Applied Research in Quality of Life*, 1(2), 151-157. <http://dx.doi.org/10.1007/s11482-006-9007-x>
- Diener, E. (2008). Myths in the science of happiness and directions for future research. In M. Eid & R. J. Larsen (Eds.), *The science of subjective well-being* (pp. 493-514). New York, NY: Guilford Press.

- Diener, E., Lucas, R., Schimmack, U., & Helliwell, J. (2009). *Well-being for public policy*. New York, NY: Oxford University Press. <http://dx.doi.org/10.1093/acprof:oso/9780195334074.001.0001>
- 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. <http://dx.doi.org/10.1111/j.0963-7214.2004.00501001.x>
- 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. <http://dx.doi.org/10.1037/0033-2909.125.2.276>
- Dodge, R. (1930). Autobiography. In C. Murchison (Ed.), *A history of psychology in autobiography* (Vol. 1, pp. 99-121). Worcester, MA: Clark University Press. <http://dx.doi.org/10.1037/11401-004>
- Easterlin, R. A. (1974). Does economic growth improve the human lot? Some empirical evidence. In P. A. David & M. W. Reder (Eds.), *Nations and households in economic growth: Essays in honor of Moses Abramovitz* (pp. 89-125). New York, NY: Academic Press.
- Flügel, J. C. (1925). A quantitative study of feeling and emotion in everyday life. *British Journal of Psychology*, 15, 318-355.
- Frey, B. S., & Stutzer, A. (2000). Maximising happiness? *German Economic Review*, 1(2), 145-167. <http://dx.doi.org/10.1111/1468-0475.00009>
- Frey, B. S., & Stutzer, A. (2002). *Happiness and economics: How the economy and institutions affect well-being*. Princeton, NJ: Princeton University Press.
- Goldings, H. J. (1954). On the avowal and projection of happiness. *Journal of Personality*, 23(1), 30-47. <http://dx.doi.org/10.1111/j.1467-6494.1954.tb02336.x>
- Gurin, G., Veroff, J., & Feld, S. (1960). *Americans view their mental health: A nationwide interview survey*. New York, NY: Basic Books.
- Hamilton, G. V. (1929). *A research in marriage*. New York, NY: A. & C. Boni.
- Hart, H. (1940). *Chart for happiness*. New York, NY: Macmillan.
- Hartmann, G. W. (1934). Personality traits associated with variations in happiness. *The Journal of Abnormal and Social Psychology*, 29(2), 202-212. <http://dx.doi.org/10.1037/h0073108>
- Hilgard, E. R. (1957). Lewis Madison Terman: 1877-1956. *The American Journal of Psychology*, 70(3), 472-479.
- Kahneman, D., Diener, E., & Schwarz, N. (Eds.). (1999). *Well-being: The foundations of hedonic psychology*. New York, NY: Russell Sage Foundation.
- Kahneman, D., Krueger, A. B., Schkade, D., Schwarz, N., & Stone, A. (2004). Toward national well-being accounts. *The American Economic Review*, 94(2), 429-434. <http://dx.doi.org/10.1257/0002828041301713>
- Klein, J. L., & Morgan, M. S. (Eds.). (2001). *The age of economic measurement*. Annual Supplement to Volume 33, History of Political Economy. Durham, NC: Duke University Press.
- Larson, R. (1978). Thirty years of research on the subjective well-being of older Americans. *Journal of Gerontology*, 33(1), 109-125.
- Layard, P. R. G. (2005). *Happiness: Lessons from a new science*. New York, NY: Penguin Press.
- Lebo, D. (1953). Some factors said to make for happiness in old age. *Journal of Clinical Psychology*, 9(4), 385-387. [http://dx.doi.org/10.1002/1097-4679\(195310\)9:4%3c385::AID-JCLP2270090418%3e3.0.CO;2-7](http://dx.doi.org/10.1002/1097-4679(195310)9:4%3c385::AID-JCLP2270090418%3e3.0.CO;2-7)
- Levy, E. (2001). Quantification, mandated science and judgment. *Studies In History and Philosophy of Science Part A*, 32(4), 723-737. [http://dx.doi.org/10.1016/S0039-3681\(01\)00019-X](http://dx.doi.org/10.1016/S0039-3681(01)00019-X)
- Lyubomirsky, S. (2008). *The how of happiness: A scientific approach to getting the life you want*. New York, NY: Penguin Press.
- McAdams, D. P. (2001). Personality psychology. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social & behavioral sciences* (Vol. 16, pp. 11308-11313). Oxford: Pergamon. <http://dx.doi.org/10.1016/B0-08-043076-7/01644-2>
- Morgan, A. E. (1934). An attempt to measure happiness. *International Journal of Ethics*, 44(2), 236-243.
- Morgan, M. S. (2001). Making measurement instruments. In J. L. Klein & M. S. Morgan (Eds.), *The age of economic measurement*, Annual Supplement to Volume 33, History of Political Economy (pp. 235-251). Durham, NC: Duke University Press.
- Murray, H. A. (1938). *Explorations in personality: A clinical and experimental study of fifty men of college age*. New York, NY: Oxford University Press.

- Myerson, A. (1917). Eupathics: A program for mental hygiene. *The Journal of Abnormal Psychology*, 12(5), 343-347. <http://dx.doi.org/10.1037/h0073646>
- Neugarten, B. L., Havighurst, R. J., & Tobin, S. S. (1961). The measurement of life satisfaction. *Journal of Gerontology*, 16(2), 134-143.
- Pollak, O. (1948). *Social adjustment in old age: A research planning report*. New York, NY: Social Science Research Council.
- Porter, T. M. (1995). *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton, NJ: Princeton University Press.
- Praag, B. M. S. V. (1968). *Individual welfare functions and consumer behavior: A theory of rational irrationality*. Amsterdam: North-Holland Pub. Co.
- Redelmeier, D. A., & Kahneman, D. (1996). Patients' memories of painful medical treatments: Real-time and retrospective evaluations of two minimally invasive procedures. *Pain*, 66(1), 3-8. [http://dx.doi.org/10.1016/0304-3959\(96\)02994-6](http://dx.doi.org/10.1016/0304-3959(96)02994-6)
- Rose, A. M. (1955). Factors Associated with the life satisfaction of middle-class, middle-aged persons. *Marriage and Family Living*, 17(1), 15-19. <http://dx.doi.org/10.2307/346773>
- Sailer, R. C. (1931). *Happiness self-estimates of young men*. Teachers College, Columbia University Contributions to Education, No. 467. New York, NY: Columbia University.
- Seligman, M. E. P. (2002). *Authentic happiness: Using the new positive psychology to realize your potential for lasting fulfillment*. New York, NY: The Free Press.
- Seligman, M. E. P., & Csikszentmihalyi, M. (2000). Positive psychology: An introduction. *American Psychologist*, 55(1), 5-14. <http://dx.doi.org/10.1037/0003-066X.55.1.5>
- Sen, A. (2008). The economics of happiness and capability. In L. Bruni, F. Comim, & M. Pugno (Eds.), *Capabilities and happiness* (pp. 16-27). Oxford: Oxford University Press.
- Symonds, P. M. (1937). Happiness as related to problems and interests. *Journal of Educational Psychology*, 28(4), 290-294. <http://dx.doi.org/10.1037/h0061067>
- Terman, L. M. (1938). *Psychological factors in marital happiness*. New York, NY: McGraw-Hill.
- Watson, G. (1930). Happiness among adult students of education. *Journal of Educational Psychology*, 21(2), 79-109. <http://dx.doi.org/10.1037/h0070539>
- Wilson, W. (1967). Correlates of avowed happiness. *Psychological Bulletin*, 67(4), 294-306. <http://dx.doi.org/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.
- Wohlgemuth, A. (1919). *Pleasure-unpleasure: An experimental investigation on the feeling-elements*. The British Journal of Psychology Monograph Supplements (Vol. 2). Cambridge: Cambridge University Press.

Appendix A



Timeline of early contributors to the literature on subjective measures of wellbeing and the science of happiness.