

NBER WORKING PAPER SERIES

FROM HAPPINESS DATA TO ECONOMIC CONCLUSIONS

Daniel J. Benjamin  
Kristen Cooper  
Ori Heffetz  
Miles S. Kimball

Working Paper 31727  
<http://www.nber.org/papers/w31727>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
September 2023

When citing this paper, please use the following: Benjamin, Daniel J., Kristen B. Cooper, Ori Heffetz, and Miles S. Kimball. 2023. "From Happiness Data to Economic Conclusions." *Annu. Rev. Econ.* 3: Submitted. DOI: <https://doi.org/10.1146/annurev-economics-081623-021136>. We are grateful to Marc Fleurbaey for helpful comments and to Colby Chambers, Dimitriy Leksanov, and Jeffrey Ohl for research assistance. For financial support, we are grateful to NIH/NIA grant R01-AG065364 to Hebrew University. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or other funding bodies. The authors have no material financial interests that relate to the research described in this paper. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2023 by Daniel J. Benjamin, Kristen Cooper, Ori Heffetz, and Miles S. Kimball. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

From Happiness Data to Economic Conclusions

Daniel J. Benjamin, Kristen Cooper, Ori Heffetz, and Miles S. Kimball

NBER Working Paper No. 31727

September 2023

JEL No. D60,D63,D9,I31

### **ABSTRACT**

Happiness data—survey respondents’ self-reported well-being (SWB)—have become increasingly common in economics research, with recent calls to use them in policymaking. Researchers have used SWB data in novel ways, for example to learn about welfare or preferences when choice data are unavailable or difficult to interpret. Focusing on leading examples of this pioneering research, the first part of this review uses a simple theoretical framework to reverse-engineer some of the crucial assumptions that underlie existing applications. The second part discusses evidence bearing on these assumptions and provides practical advice to the agencies and institutions that generate SWB data, the researchers who use them, and the policymakers who may use the resulting research. While we advocate creative uses of SWB data in economics, we caution that their use in policy will likely require both additional data collection and further research to better understand the data.

Daniel J. Benjamin  
University of California Los Angeles  
Anderson School of Management  
and David Geffen School of Medicine  
110 Westwood Plaza  
Entrepreneurs Hall Suite C515  
Los Angeles, CA 90095  
and NBER  
daniel.benjamin@anderson.ucla.edu

Kristen Cooper  
Gordon College  
229 Jenks  
255 Grapevine Road  
Wenham, MA 01984  
kristen.cooper@gordon.edu

Ori Heffetz  
S.C. Johnson Graduate School of Management  
Cornell University  
324 Sage Hall  
Ithaca, NY 14853  
and The Hebrew University of Jerusalem  
and also NBER  
oh33@cornell.edu

Miles S. Kimball  
Economics Building  
University of Colorado Boulder  
261 UCB  
#212  
Boulder, CO 80302  
and NBER  
miles.kimball@colorado.edu

Self-reported well-being (SWB) data have become increasingly common in economics research. Informally referred to as “happiness data,” they are based on survey questions that ask respondents about their happiness, life satisfaction, or other emotions or other life evaluations. Over the past half century, researchers in this field have brought SWB data to bear on some of the most important and difficult questions in economics, from exploring the welfare consequences of economic growth (Easterlin, 1974) to examining life-cycle theory and policy (Deaton, 2018). Indeed, SWB data have been advocated as a direct measure of preferences and welfare, and as especially valuable in contexts where choice data are unavailable or may not reliably reflect preferences.

There have been many excellent reviews of the SWB literature in economics.<sup>1</sup> These reviews mainly take a bird’s-eye view, focusing on main empirical findings and often also covering, at least briefly, the literature’s normative basis, econometric methods, and remaining challenges. Departing from, and complementing, those and other like reviews, this review takes an opposite, low-flying approach. We zoom in on, and explore in detail, several central assumptions that lie at the heart of the use of SWB data in economics, and which must underlie any consistent, theory-guided interpretation of SWB applications in economics. We do so in two steps. In the first—Sections I–III—we use a simple theoretical framework to reverse-engineer some of the crucial assumptions that underlie leading applications. In the second step—Sections IV–VI—we discuss evidence bearing on these assumptions and provide practical advice to the agencies and institutions that generate SWB data, the researchers who use the resulting data in research, and the policymakers who seek to use the resulting research to make better policy—now and in the future.

We aim to convey two reactions to the literature: excitement and caution. We are excited about both the creative use of SWB data by “happiness” economists and the openness to SWB-based evidence by more mainstream economists—an openness that reflects, more generally, a welcome relaxation of the historical taboo on non-choice data. Yet we caution that without

---

<sup>1</sup> These include Frey and Stutzer (2002), Di Tella and MacCulloch (2006), Stutzer and Frey (2010), Helliwell and Barrington-Leigh (2010), Van Praag and Ferrer-i-Carbonell (2010), MacKerron (2012), Piekalkiewicz (2017), Lane (2017), and Clark (2018), as well as the 2012-2023 *World Happiness Reports* (e.g., WHR, 2023), a number of chapters in the *Oxford Handbook of Well-Being and Public Policy* (Fujiwara and Dolan, 2016; Graham, 2016; Clark, 2016; Decancq and Neumann, 2016) and the *Handbook of Labor, Human Resources and Population Economics* (Nikolova and Graham, 2021; Ifcher, Zarghamee, and Goff, 2021; Easterlin and O’Connor, 2022; Paul Frijters, 2021; and others), and Layard and De Neve’s (2023) textbook.

transparent, empirically supported assumptions, the resulting findings are hard to interpret, and their potential use in policy is limited.

We start with preliminaries in Sections I and II. To make concrete what we mean by *SWB data*, in Section I we list some of the large-scale surveys that economists have been using as sources for such data. For each, we provide the exact language used in its main SWB survey question or questions, list the available response options, and provide survey details including geographic coverage and survey response rate. We also discuss the promise of these data, and their potential advantages (over choice data alone) as welfare indicators.

To make concrete what such data may capture conceptually, in Section II we present a simple theoretical framework for incorporating SWB data into economic models. Crucially, we distinguish between the empirical measures—the *SWB data*—and the underlying theoretical and philosophical concept they aim to measure—the *well-being* (or *welfare*, or *utility*) of certain individuals during a certain period of time. Depending on one’s normative views and interpretation of the data, the latter may be considered the ultimate objective of policy, or an important component of it. In the standard economic approach, the normative welfare criterion is preference satisfaction. Under this view, SWB measures are appealing because they may inform us about preferences. Our theoretical framework can also accommodate alternative normative views.

We close our discussion of preliminaries by discussing the types of assumptions relating SWB and utility that must underlie the use and interpretation of SWB data in economic applications. We also observe that in order to link the SWB measures from Section I to the theory from Section II, additional strong survey-methodology assumptions are required. For example, almost all SWB applications have to rule out systematic effects of survey-question order, survey-nonresponse bias, and respondent-level response style. Moreover, *all* SWB applications have to make additional strong *conceptual* assumptions—for example, assuming that well-being itself is interpersonally comparable.

Having set the stage, Section III is the heart of the first part of our review. We focus on leading applications of SWB data in economics, and discuss the assumptions that they *actually* make, if implicitly, about what the data represent. Using a theoretical framework of time-separable utility, we clarify the theoretical assumptions underlying recent applications regarding which time horizon is captured by the SWB data. We also briefly discuss assumptions about

which social dimension SWB data capture. We document widespread inconsistency of assumptions across papers and across applications—in some cases, with different assumptions made by different researchers about the very same data. For example, assumptions vary about whether a specific SWB question captures flow (i.e., instantaneous) utility, or lifetime utility, or the forward-looking expected present discounted value of utility flows.

We also show in Section III how the appropriate scientific and policy conclusions that can be drawn in typical applications often hinge on which, if any, theoretical assumption holds. For example, researchers often run a regression of current SWB on current income and current consumption of a nonmarket good such as employment (e.g., Clark and Oswald, 2002), and calculate the money value of the good as the ratio of coefficients. We show that the theoretical interpretation of this empirical ratio depends on which utility notion is captured by SWB. In particular, if SWB measures flow utility, then the estimated “dollar cost of unemployment” is a one-time cost, whereas if SWB measures lifetime utility, then (under some simplifying assumptions) it is an *annual* cost—so the total cost will be many times higher. We also illustrate, for example, how assumptions about the SWB measure matter crucially for interpreting SWB’s increase with age later in life (e.g., Blanchflower and Oswald, 2008; Stone, Schwartz, Broderick, and Deaton, 2010).

In our review’s second part, Sections IV–VI, we evaluate these assumptions in light of empirical evidence to date. Drawing on our own research as well as on others’, we make concrete proposals regarding what we believe are better empirically supported assumptions and (hence) interpretations of existing SWB measures (in Section IV), better approaches for addressing response-scale and scale-use issues in SWB data (in Section V), and better collection of new SWB measures, guided by theory and evidence (in Section VI).

Specifically, in Section IV we synthesize a list of insights for the researchers who use existing SWB data such as those from the surveys in Section I, as well as for consumers of the research, such as policymakers. We also give advice to those who collect SWB data. We start with a warning regarding survey methodology: evidence suggests that SWB data may be particularly subject to survey- and sample-design issues in ways that may affect, and even reverse, researchers’ conclusions. Our advice is to diagnose the problem with simple tools that could help assess the extent of the problem in specific applications. Researchers should demand and use, for example, paradata that are already collected by survey centers but are not yet widely

shared. For another example, the designers of surveys should more often use randomization to assess the effects of survey-design decisions. We close the section with a quick list of proposals regarding the conditions under which, in our reading of the evidence, we think that SWB provides better or worse measures of preferences. We focus on the choice setting, the choice options, and the SWB measure.

In Section V, we discuss in detail issues related to heterogeneity in response-scale use and other response-scale issues for SWB data. We review past research on this topic, and make concrete proposals regarding both interpretation of existing data and collection of new auxiliary data that would allow researchers to better deal with these issues in the future.

In Section VI we take the next step towards developing new SWB measures that are better guided by both theory and evidence. We present evidence and give guidance both on how to design new SWB questions that may come closer to capturing specific utility notions that researchers are interested in and on how to aggregate SWB questions that capture different aspects of well-being into a single, individual-level SWB index.

We conclude in Section VII, where we provide a table that can serve as a quick reference summarizing the advice in this review (by section). We hope that other researchers can build on and expand this table, adding their own advice. Such organized, accumulated knowledge would help all three key groups: generators of SWB data, users of such data, and consumers of the resulting research.

## **I. SWB Measures and Their Promise**

### **I.A. SWB measures and surveys**

Angner (2011) reviews the historical roots of SWB measures, originally designed between the 1920s and 1970s. He notes that the SWB questions routinely used nowadays by economists were originally designed by *non*-economists—marriage researchers, education and personality psychologists, mental-health epidemiologists, gerontologists, and social-indicator researchers—for a variety of purposes that did *not* include economics research. Thus, there is no particular reason why these questions should capture the specific utility notions that present-day economists are interested in.

Some of these SWB questions have been collected for many decades, using reasonably comparable survey designs, on somewhat comparable populations, in a variety of regions, states,

or countries. This has made them attractive to economics researchers, in spite of their weaknesses. Indeed, empirical economists, who are accustomed to trading off data quality for data availability, routinely find creative ways to exploit less-than-perfect data; SWB data are no exception.

Table 1 provides details and lists the exact wording of SWB questions from seven current large-scale surveys that are commonly used by economists as sources of SWB data: questions about life satisfaction (in the BRFSS, Eurobarometer, and ESS), happiness these days (Eurobarometer and GSS), evaluation of life in terms of steps on a ladder (GSWBI and Gallup World Poll), happiness and other emotions in the past week (HRS), and a combination of questions (which we informally refer to as “The UK Four”) that includes life satisfaction, things in life being worthwhile, happiness yesterday, and anxiety yesterday (ONS).<sup>2</sup>

Let us make a few observations. Data go back at least a decade and, for the GSS and Eurobarometer, five decades. Geographic coverage prominently includes the US, UK, and EU. The surveys all use discrete response scales for the self-reported levels, with a range from only two (Yes/No) to eleven (0–10) response options. Finally, recent survey response rates, having long been in decline, are typically below 50 percent, and often much below. We return to the challenges associated with such response scales and response rates in later sections.

=== TABLE 1 AROUND HERE (see next page). ===

---

<sup>2</sup> Different SWB questions may be hypothesized to be more compatible with different philosophical views (discussed in Section II below). In particular, a happiness question may be aimed at measuring the objective in a hedonistic view, whereas life satisfaction may be aimed at measuring the objective in a preference-satisfaction view.

**Table 1: Surveys and SWB Questions Commonly Used by Economists**

<u>Survey name</u> . Year of first wave. Geographic coverage. Response rate in recent year(s).
<i>SWB question(s)</i>
Response options and response scale(s).
<p><u>Behavioral Risk Factor Surveillance System</u> (BRFSS). 1984. US. 44% (state/territory/DC median) in 2021.</p> <p><i>In general, how satisfied are you with your life?</i></p> <p>4 options: Very satisfied / Satisfied / Dissatisfied / Very dissatisfied.</p>
<p><u>Eurobarometer</u>. 1974. EU countries. 45% (27-country median; range: 24–68%) in 2023.</p> <p><i>Taking all things together, how would you say things are these days—would you say you’re very happy, fairly happy, or not too happy these days?</i></p> <p><i>On the whole, are you very satisfied, fairly satisfied, not very satisfied or not all satisfied with the life you lead?</i></p> <p>3 options: Very happy / Fairly happy / Not too happy.</p> <p>4 options: Very satisfied / Fairly satisfied / Not very satisfied / Not at all satisfied.</p>
<p><u>European Social Survey</u> (ESS). 2002. European countries. 39% (31-country median; range: 15–73%) in 2020–2022.</p> <p><i>All things considered, how satisfied are you with your life as a whole nowadays?</i></p> <p><i>Taking all things together, how happy would you say you are?</i></p> <p>11 options: 0–10 (Extremely dissatisfied/unhappy–Extremely satisfied/happy).</p>
<p><u>Gallup-Sharecare [was: -Healthways] Well-Being Index</u> (GSWBI [was: GHWBI]). 2008. US, UK. 11% (estimate*).</p> <p><i>Please imagine a ladder with steps numbered from 0 at the bottom to 10 at the top. The top of the ladder represents the best possible life for you and the bottom of the ladder represents the worst possible life for you.</i></p> <p><i>On which step of the ladder would you say you personally feel you stand at this time?</i></p> <p>11 options: 0–10.</p>
<p><u>Gallup World Poll</u>. 2005. 142 countries in 2022. 14–69% (rich-country–poorer-country median) in 2006–2016**.</p> <p>Same SWB question and response options as in the GSWBI.</p>
<p><u>General Social Survey</u> (GSS). 1972. US. 51% in 2022.</p> <p><i>Taken all together, how would you say things are these days--would you say that you are very happy, pretty happy or not too happy?</i></p> <p>3 options: Very happy / Pretty happy / Not too happy.</p>
<p><u>Health and Retirement Study</u> (HRS). 1992. US. Differently defined response rates (longitudinal panel study).</p> <p><i>Now think about the past week and the feelings you have experienced. Please tell us if each of the following was true for you much of the time during the past week. ... You felt you were happy. / You felt sad. / You enjoyed life. / You felt depressed.</i></p> <p>2 options: Yes / No.</p>
<p><u>Office for National Statistics</u> (ONS) Well-being. 2011. UK. 17% in Jan–Mar 2023.</p> <p><i>Overall, how satisfied are you with your life nowadays?</i></p> <p><i>Overall, to what extent do you feel that the things you do in your life are worthwhile?</i></p> <p><i>Overall, how happy did you feel yesterday?</i></p> <p><i>Overall, how anxious did you feel yesterday?</i></p> <p>11 options: 0–10 (e.g., Not at all anxious–Completely anxious.)</p>

**Notes:** For sources, sample sizes, and other information, see Supplementary Material.

\* Reported in Skopec, Musco, and Sommers (2014) as “approximate” response rate from “Gallup-Healthways estimate of daily tracking poll response rate.” No specific year is mentioned. Of potential concern, Gallup does not officially disclose response rates externally (based on our communications with Gallup, August 2023).

\*\* Reported in Deaton (2018). Gallup does not officially disclose response rates externally (see previous table note).



In the rest of this review we focus on applications that use SWB data from questions such as those in Table 1. For completeness, we note here that almost three decades ago, Kahneman, Wakker, and Sarin (1997) advocated an alternative to such single-question SWB measures: continuous measurement of current feelings of happiness or pain. Two decades ago, a group of prominent researchers from psychology, economics, and other fields (Kahneman, Krueger, Schkade, Schwarz, and Stone, 2004) proposed a less burdensome approximation to such continuous measurement. Their proposed “Day Reconstruction Method” (DRM) involved respondents filling out an hourly log where they reconstruct, at the end of the day, their activities during different parts of the day, along with details such as whom they were with and their states of mind during each activity. That method has not been widely used by economists. One possible reason is that such data, which are costly to collect, are not available in large-scale surveys. Another is that economists have been mainly interested in the effects of variables that vary at much-lower-than-daily frequency, such as income and unemployment.

## **I.B. SWB measures and preferences**

What underlying concepts, of interest to economists, could an SWB measure capture? A *self-reported*-well-being measure is a respondent’s answer to a survey question. It is conceptually distinct from the respondent’s *underlying* well-being, which is the quality of the person’s life as assessed according to some philosophical view of what makes life good.<sup>3</sup> However, SWB may be *informative* about well-being—indeed, this possibility is the main reason that SWB could be a valuable source of data to economists and policymakers.

Our theoretical framework in the next section is grounded in neoclassical welfare economics, which is based on the philosophical view that a person’s well-being is related to the satisfaction of that person’s preferences (e.g., Hausman, 2012). Our framework is thus focused on what information about preference satisfaction is provided by SWB.

In neoclassical welfare economics, an individual’s preferences are operationalized starting from the principle of revealed preference: preferences correspond to what the individual

---

<sup>3</sup> In our experience, a common source of conceptual slippage is semantic. For example, the same word “happiness” is used to describe a feeling, the answer to a survey question about happiness, and the ultimate criterion that people aim to maximize (indeed, economics textbooks often use the words “utility” and “happiness” interchangeably). Similarly, the phrase “life satisfaction” is used to describe both satisfaction of preferences about one’s life and the answer to a survey question about life satisfaction.

would choose under ideal conditions. Depending on the specific normative theory, these ideal conditions may include some or all of the following: full information and correct beliefs about the consequences of different options; time, ability, and motivation to fully think through the consequences; and absence of temptation and other behavioral distortions (e.g., Railton 1986; Goodin 1986; Hausman, 2012; Bernheim, 2016). We refer to choices made under such conditions as *idealized choices*. In practice, idealized choices are never observed, and economists instead treat choices as more reliable indicators of preferences if they more closely approximate idealized choices—for example, high-stakes choices made repeatedly with extensive feedback regarding non-tempting consequences (Beshears, Choi, Laibson, and Madrian, 2008).

Choices could have four potential disadvantages, relative to SWB, as a measure of (or a source of information about) preferences: choices are made under non-ideal conditions, adequate choice data may be unavailable either in principle or in practice, and choices (and choice-related proxies) may provide too little information. We consider the flipside of each in turn, highlighting the promise of SWB data.

First, there is a fundamental difference between choices and SWB: A choice is a ranking of options *ex ante*, i.e., prior to experiencing the consequences and based on beliefs, whereas SWB is a reported evaluation of a situation that is more, or entirely, *ex post*, i.e., after having experienced at least some of the consequences (although SWB may still incorporate expectations regarding additional consequences further in the future). SWB can thus provide valuable data on well-being in settings where choices are an unreliable guide to preferences. Thus, as SWB researchers have highlighted, SWB can be especially useful in contexts where learning, adaptation, changes in budget constraints, or changes in beliefs and preferences are difficult to anticipate (e.g., Kahneman, Wakker, and Sarin, 1997; Dolan and Kahneman, 2008). Another common use of SWB data is to learn about preferences in contexts where observed choices may be subject to behavioral biases (to a greater extent than SWB), for example, preferences over cigarette smoking (Gruber and Mullainathan, 2005).

Second, key data about the relevant choices may be unavailable. In particular, in trying to gauge lifetime utility, many of the relevant choices are in the past, and only the final choice may be known (if that), not what alternative options and information the individual had at the time.

Third, SWB data may be available in contexts where there are no relevant choices to study. For example, researchers often want to compare well-being across countries (Deaton,

2008) or across men and women (Stevenson and Wolfers, 2009). While we cannot observe people's choices of where, or into what sex, to be born (and even if we did, such choices by the yet-unborn would be subject to the concerns from the previous two paragraphs), SWB measures that elicit flow, forward-looking or lifetime utility can at least be imagined (and possibly constructed). Other examples include studying the effect on well-being of an event beyond an individual's control, such as death of a family member (e.g., Deaton, Fortson, and Tortora, 2010; Kimball, Nunn, Silverman, 2015), or the effect of aging (e.g., Stone, Schwartz, Broderick, and Deaton, 2010).

Finally, for many investigations, SWB may provide a more direct gauge of well-being (requiring fewer assumptions for inference) than choices. Consider a study of the effect of income on well-being that uses exogenous variation in lottery winnings (e.g., Lindqvist, Östling, and Cesarini, 2020). In principle, it might be possible to learn about preferences from people's lottery-participation choices, but the information is sparse, and many auxiliary assumptions would be required (e.g., about probability beliefs and opportunity costs). Directly comparing SWB of winners and non-winners is much more straightforward. As another example, consider the effect on well-being of the Moving to Opportunity experiment (Ludwig et al., 2012). While the effect on overall well-being could in principle be estimated by examining effects on various outcomes and assigning valuations to them, SWB might provide a more comprehensive measure of well-being that reflects each individual's idiosyncratic, subjective valuations.

## **II. SWB Measures in Theory**

### **II.A. Theoretical framework**

We adapt a theoretical framework from Benjamin, Fleurbaey, Guzman Debnam, Heffetz, and Kimball (forthcoming)<sup>4</sup> to describe the main assumptions that economists have effectively made (if only implicitly) about the relationship between SWB and well-being. As mentioned in the previous section, this framework is focused on what information about preference satisfaction is provided by SWB. However, it can accommodate alternative philosophical views. In particular, prominent advocates of SWB as a goal of policy sometimes do so based on a

---

<sup>4</sup> Relative to Benjamin, Guzman Debnam, Fleurbaey, Heffetz, and Kimball (forthcoming), our framework is simplified because we omit the explicit description of other-regarding preferences (treated here as components of the consumption vectors) and less general in that we assume exponential discounting. At the same time, unlike the earlier paper, we make explicit that the SWB report will vary depending on the SWB question and response scale.

hedonistic theory of well-being (e.g., Layard, 2005) that traces back to Jeremy Bentham’s (1789) utilitarianism and classical economics. According to this philosophical view, well-being consists of a person’s happiness (net of pain). Since *self-reported* happiness (a response to a survey question) is conceptually different from the respondent’s *underlying* happiness (the welfare criterion), our framework still applies under such a hedonistic theory of well-being, with the words “preferences” and “utility” below replaced by “underlying happiness.”<sup>5</sup>

As we will discuss in Section III, papers in the SWB literature often assume that SWB data capture only certain components of preferences. To facilitate defining these components, we assume that preferences can be represented by discounted expected utility with exponential discounting.<sup>6</sup> We denote an individual’s lifetime sequence of consumption vectors by  $\mathbf{c}_0, \mathbf{c}_1, \dots, \mathbf{c}_T$ . Without loss of generality, the consumption vectors include not only everything that affects preferences (including both market and non-market goods) but also everything that affects SWB. The individual’s *lifetime utility* in any period  $t$ , given their (possibly mistaken) information and beliefs, is:

$$(1) \quad U_t^{\text{lifetime}} = E_t \sum_{\tau=0}^T \delta^{\tau-t} u(\mathbf{c}_\tau),$$

where we refer to  $u(\mathbf{c}_t)$  as the individual’s *flow utility* in period  $t$ .

Even idealized choices do not identify lifetime utility but instead only identify *forward-looking utility*:<sup>7</sup>

---

<sup>5</sup> In addition to preference satisfaction and hedonism, philosophers’ categorization of views about well-being include a third type, called an objective-list view. Within economics, the best-known such view is Amartya Sen’s (1985) and Martha Nussbaum’s (2000) capability approach, according to which (regardless of the person’s preferences) a person’s well-being consists of having all the necessary freedoms to achieve certain “doings” (activities a person could undertake) and “beings” (the kind of person someone could be). Under this view, neither utility nor SWB responses are reliable measures of well-being, although they may be informative about well-being.

<sup>6</sup> We assume expected utility and exponential discounting for concreteness and simplicity, but nothing in our framework depends on it. For example, we could instead assume probability weighting or present-biased preferences. The important features of the framework are the dependence of preferences on the individual’s information and beliefs in period  $t$ , which we discuss in Section III.A, and time separability, which is needed for Assumption 1 below.

<sup>7</sup> If preferences are not time-separable—for example, if individuals are averse to anticipated regret (Loomes and Sugden, 1982) and if anticipated regret is truly part of preferences rather than being a bias—then idealized choices can identify some components of past consumption. For our discussion here, the important point is that idealized choices do not identify lifetime utility.

$$(2) \quad U_t^{\text{forward}} = E_t \sum_{\tau=t}^T \delta^{\tau-t} u(\mathbf{c}_\tau),$$

with the summation beginning in period  $t$  instead of period 0. This is because for a choice made in period  $t$ , the consumption vectors in prior periods are constant across the options.

## II.B. A Taxonomy of Assumptions

The potential advantages of SWB data discussed in Section I open exciting avenues of economic research on important topics. But a specific application's validity and interpretation depend on what information about preferences or welfare is captured by SWB data. In this review we detail several fundamental types of assumptions: on the time horizon about which SWB data provide information (Assumption 1); on the comprehensiveness of SWB data (also Assumption 1); on interpersonal and intergroup comparability (Assumptions 2, 3 and 4); and on the survey and sample the SWB data originate from (not numbered).

Time-horizon assumptions typically come in one of three alternative variants (1A, 1B and 1C):

**Assumption 1A.** SWB is ordinally equivalent to lifetime utility  $U_t^{\text{lifetime}}$ .

**Assumption 1B.** SWB is ordinally equivalent to forward-looking utility  $U_t^{\text{forward}}$ .

**Assumption 1C.** SWB is ordinally equivalent to flow utility  $u(\mathbf{c}_t)$ .

Other variants of Assumption 1 are sometimes mentioned or implied by the literature, but they do not motivate SWB applications as 1A–1C do. For example, some SWB questions ask explicitly about happiness or other emotions in the recent past (e.g., questions in the HRS, GSWBI, and ONS in Table 1). The use of these questions could in principle be motivated by a backward-looking utility notion, for example, as a building block for a social objective based on lifetime utilities. However, most applications use them instead to measure recently experienced flow utility.

There are a few examples of papers making an alternative assumption that happiness is, like other emotions, an argument of the utility function (Kimball and Willis, 2006; Becker and

Rayo, 2008; Benjamin, Heffetz, Kimball, and Rees-Jones, 2012)—but always (as far as we are aware, and at least implicitly) in the context of considering one of the above variants of Assumption 1. Similarly, as we discuss in Section IV.B below, many researchers nowadays appear to assume that different SWB measures capture different aspects of well-being, i.e., in our framework, period  $t$ 's SWB (by some SWB measure) does not necessarily capture  $u(c_t)$ , and could instead capture some other  $\tilde{u}(c_t)$  (with different implied marginal rates of substitution)—but, again, such an assumption is made in the context of discussing some variant of Assumption 1.

Assumption 1 would be sufficient for using SWB as a measure of an individual's ordinal preferences. However, applications using SWB data nearly always also draw conclusions about well-being by comparing SWB across individuals—specifically, comparing mean SWB across groups of individuals (often in a regression framework, with controls). This practice requires additional, stronger assumptions involving cardinality, interpersonal comparisons, and group well-being. We enumerate these as three implicit, conceptually distinct assumptions:

**Assumption 2.** Well-being is interpersonally comparable in both levels and differences.

**Assumption 3.** SWB is ordinally equivalent to well-being, and the monotonic transformation from well-being to SWB is the same across individuals. That is, people use the SWB response scale in the same way when *reporting* their well-being.

**Assumption 4.** Well-being, as measured on the scale of SWB reports, can be averaged (or added up) to obtain a valid measure of a group's well-being.

The notion of a group's well-being requires a social welfare function, and Assumption 2 is a prerequisite for defining a social welfare function (see, e.g., Adler, 2019, chapter 2). Assumption 3 is needed in order to use SWB as a measure of well-being. Moreover, even if well-being is interpersonally comparable (Assumption 2), SWB cannot be interpersonally comparable unless the monotonic transformation from well-being to SWB is the same across individuals; Assumption 3 makes this premise explicit. Assumption 4 states that SWB is measured on an appropriate scale such that mean SWB is a valid social welfare function.

Finally, any application of SWB data requires strong survey-methodology assumptions. For SWB responses to be stable and reproducible, and for results to generalize beyond the survey

sample to a broader population of interest, researchers have to rule out systematic effects of survey-question order, sample self-selection, and survey-nonresponse bias, to name just a few assumptions.

### **III. SWB Applications and their Underlying Assumptions**

All applications discussed in this review make Assumptions 2–4 and the additional survey-methodology assumptions from the previous section—as do virtually all SWB applications in economics that we are aware of. (We discuss evidence on the survey-methodology assumptions in Section IV, and on Assumptions 3 and 4 in Section V.) But which variant of Assumption 1—lifetime utility (1A), forward-looking utility (1B), or the narrower flow utility (1C)—is assumed by different SWB applications? Does it matter? In this section we answer these two questions in the context of several important applications. We also briefly discuss an orthogonal but analogous issue—What other-regarding preferences are captured by responses to the SWB question?—that is not explicitly modeled by our theoretical framework, but that could be easily added (see footnote 4).

#### **III.A Time horizon**

Few papers that apply SWB data explicitly discuss which intertemporal preference information is captured by the SWB measure. Exceptions include Gruber and Mullainathan (2005) and Blanchflower and Oswald (2004), both of whom analyze the GSS happiness question (see Table 1 for this and all other SWB questions discussed in this section). In their 2002 working paper, Gruber and Mullainathan (pp. 24, 28–29) argue that their evidence is most consistent with forward-looking utility. In contrast, Blanchflower and Oswald (p. 1362) state that the *same* question is “more naturally interpreted as a flow rather than a stock.” Another exception is Alesina, Di Tella, and MacCulloch (2004), who also analyze this same GSS happiness question, as well as the life satisfaction question from the Eurobarometer Survey. Referencing Kahneman, Wakker, and Sarin’s (1997) concept of “experienced utility” as a flow of affective experiences,<sup>8</sup> Alesina et al. appear to interpret their SWB questions as capturing flow

---

<sup>8</sup> Kahneman, Wakker, and Sarin themselves are not only explicit about assuming that their SWB question is measuring “instant utility,” but they also use SWB questions that ask a respondent to answer about the immediate present—an example of consistency between assumption and SWB question that we argue should be followed more generally. However, the specific SWB questions they study, such as “rate the current intensity of pain” prompted

utility: “Our paper, and we believe much of the happiness literature, can be understood as an application of experienced utility, a concept that emphasizes the pleasures derived from consumption” (their footnote 7). Yet later, after finding that greater national inequality reduces individuals’ SWB, they argue that a plausible mechanism is beliefs about how national inequality affect one’s own future prospects—a mechanism that is more naturally interpreted under the assumption that SWB captures forward-looking or lifetime utility.

The other two papers we are aware of that make explicit statements about their time-horizon assumptions are Finkelstein, Luttmer, and Notowidigdo (2013) and Aghion, Akcigit, Deaton, and Roulet (2016). Finkelstein et al. (2013) study the HRS happiness question and write: “As is standard in the happiness literature, we interpret the happiness question as a proxy for von Neumann–Morgenstern (flow) utility” (p. 234). Aghion et al. (2016) treat the GHWBI ladder question and the BRFSS life satisfaction question as measures of forward-looking utility: “Life satisfaction is captured by the expected discounted valuation of an individual’s future earnings” (p. 3870).

In most applications of SWB data, including the many in which survey respondents’ time-horizon interpretation of the SWB question is not discussed, it matters for the conclusions that can be drawn. Here we give three examples.

Our first example is common in economic applications: using SWB data to calculate the money valuation of a good. Specifically, researchers run a regression of current SWB on current income  $y_t$  and current consumption of a “good”  $x_t$ , and calculate the money value of the good as the ratio of coefficients. The appropriate interpretation of the resulting value depends on what preference information is captured by the SWB data. For example, if SWB measures flow utility  $u(c_t) \equiv u_t$ , then this coefficient ratio is  $\text{ratio}_u \equiv \frac{\text{Cov}(u_t, \check{x}_t)/\text{Var}(\check{x}_t)}{\text{Cov}(u_t, \check{y}_t)/\text{Var}(\check{y}_t)}$ , where  $\check{x}_t$  is the residual from a regression of  $x_t$  on all covariates (including  $y_t$ ) and  $\check{y}_t$  is the residual from a regression of  $y_t$  on all covariates (including  $x_t$ ). If SWB measures a broader notion of utility  $U_t$ , such as lifetime or forward-looking utility, then the coefficient ratio is  $\text{ratio}_U \equiv \frac{\text{Cov}(U_t, \check{x}_t)/\text{Var}(\check{x}_t)}{\text{Cov}(U_t, \check{y}_t)/\text{Var}(\check{y}_t)}$ . These ratios are equal if and only if  $\frac{\text{Cov}(U_t - u_t, \check{x}_t)}{\text{Cov}(u_t, \check{x}_t)} = \frac{\text{Cov}(U_t - u_t, \check{y}_t)}{\text{Cov}(u_t, \check{y}_t)}$ . In words, current  $\check{x}_t$  and  $\check{y}_t$  must have the same ratio of covariance with non-current utility,  $U_t - u_t$ , to covariance with current utility,  $u_t$ . This

---

every 60 seconds (Redelmeier and Kahneman, 1996), have not been widely used in the SWB literature and so are not a focus of this paper, as we discussed in Section I.



condition would hold if, for example, the regression coefficients were identified by exogenous shocks in  $\tilde{x}_t$  and  $\tilde{y}_t$  that affect flow utility only in period  $t$ . However, in practice the condition is unlikely to hold because in nearly all applications, the regression coefficients are identified by cross-sectional variation, and  $\tilde{x}_t$  and  $\tilde{y}_t$  typically have different time profiles of covariance with flow utility.

We illustrate with the example of calculating the money valuation of unemployment status, a common application in the SWB literature (e.g., Blanchflower and Oswald, 2004). Here,  $x_t$  is employment status. Consider three assumptions that dramatically simplify a more complicated reality but help make the point clear: (a) unemployment conditional on the covariates occurs randomly, lasts one period, and only affects current flow utility; (b) the effect of current income conditional on the covariates,  $\tilde{y}_t$ , on current flow utility is identical across periods; and (c) cross-sectional variation in current income  $y_t$  fully reflects cross-sectional variation in permanent income. Under assumption (a),  $\text{Cov}(U_t, \tilde{x}_t) = \text{Cov}(u_t, \tilde{x}_t)$ , so the numerators of  $\text{ratio}_u$  and  $\text{ratio}_U$  are equal. However, under assumptions (b) and (c), the denominators of  $\text{ratio}_u$  and  $\text{ratio}_U$  differ substantially; for example, if  $U_t$  is forward-looking utility and  $t < T$ , then  $\text{Cov}(U_t, \tilde{y}_t) = \sum_{\tau=t}^T \delta^{\tau-t} \text{Cov}(u_\tau, \tilde{y}_\tau) \gg \text{Cov}(u_t, \tilde{y}_t)$ . In words,  $\text{ratio}_u$  is an estimator of the “dollar cost of unemployment” in units of current income—that is, a one-time cost—whereas  $\text{ratio}_U$  is an estimator of this cost in units of permanent income—that is, an annually recurring cost. Thus, if the SWB measure is assumed to capture forward-looking utility instead of flow utility, then the (same) numerical estimate from the SWB regression implies a cost of unemployment many times larger!

Even in applications where cross-sectional variation in  $x_t$  is permanent—e.g., when  $x_t$  is the death of a family member (e.g., Deaton, Fortson, and Tortora 2010)—the effects on flow utility may diminish over time due, for example, to hedonic adaptation, implying that  $\tilde{x}_t$  and  $\tilde{y}_t$  are still likely to have different time profiles of covariance with flow utility. In cases of costly investment—for example, living through difficult years of schooling in order to increase utility in future years— $\text{Cov}(u_t, \tilde{x}_t)$  and  $\text{Cov}(U_t, \tilde{x}_t)$  could even have opposite signs.

Second, another common application of SWB data, especially in the psychology literature, is to infer that people are making mistakes when a deviation from usual behavior is found to increase SWB. For example, Dunn, Aknin, and Norton (2008) find that experimental participants randomly assigned to spend money on someone else rather than on themselves were

happier when surveyed later the same day. This and related findings, together with survey evidence that people expect spending on themselves to make them happier than spending on others, lead Dunn et al. to conclude that “policy interventions that promote prosocial spending...may be worthwhile.” However, the inference that people are making a mistake is only warranted if SWB represents forward-looking utility. Otherwise, apparent “mistakes” could instead reflect optimal intertemporal tradeoffs. Indeed, consistent with this possibility, Falk and Graeber (2020) find that experimental participants randomly assigned to donate money to charity rather than receive money themselves were *less* happy four weeks later (possibly due to the foregone consumption on self), despite being happier at the end of the lab session.

Finally, there is a growing literature on how SWB varies with age. Unlike our examples above, where SWB data are used as a means to answer an economic question—e.g., to price a nonmarket good—that literature views SWB as a variable of direct interest. Much of the evidence points to a U-shape, with SWB reaching a nadir in middle age, but there is no consensus on the reason for this pattern (e.g., Blanchflower and Oswald, 2008; Blanchflower, 2020). While people might change how they use the SWB response scale with age (as suggested by Stone, Schneider, Junghaenel, and Broderick, 2019), few analyses account for this possibility. Therefore, consistent with the literature, our discussion here assumes that, with age, the utility notion captured by the SWB question, which we denote by  $\tilde{U}_t$ , remains throughout life the same monotonic transformation of the utility notion  $U_t$  that a particular paper focuses on.

In the only discussion we know of relating age to the intertemporal preference information that is captured by the SWB measure, Finkelstein et al. (2013, footnote 15) write: “[Forward-looking utility] seems inconsistent with the empirical finding that happiness increases with age for older people (unless one believes growing older means fewer future years with negative flow utility).” Using our notation, if SWB is forward-looking utility, then  $\tilde{U}_{t+1} - \tilde{U}_t = \sum_{\tau=t+1}^T [E_{t+1} \delta^{\tau-(t+1)} u(\mathbf{c}_\tau) - E_t \delta^{\tau-t} u(\mathbf{c}_\tau)] - u(\mathbf{c}_t)$ . The sign of this expression is less clear than Finkelstein et al. suggest because  $\delta^{\tau-(t+1)} > \delta^{\tau-t}$ , so the summation term may generally be expected to be positive. Finkelstein et al.’s argument is silent regarding this term, and relates only to the other term, which is only positive if  $u(\mathbf{c}_t) < 0$ .

Moreover, the interpretation of the SWB measure is relevant for what mechanisms may explain the U-shape. If it is lifetime utility, then the change in SWB from age  $t$  to  $t + 1$  is  $\tilde{U}_{t+1} - \tilde{U}_t = \sum_{\tau=0}^T [E_{t+1} \delta^{\tau-(t+1)} u(\mathbf{c}_\tau) - E_t \delta^{\tau-t} u(\mathbf{c}_\tau)]$ . In that case, changes in SWB with age

are due to unanticipated shocks to flow utility, or due to belief updates. In contrast, if SWB measures flow utility, then variation with age is unrelated to both discount factors and beliefs:  $\tilde{U}_{t+1} - \tilde{U}_t = u(c_{t+1}) - u(c_t)$ , implying that the quantity or quality of consumption in at least some domains of life increases with age. Finally, consider a tweak to Assumption 1B (SWB is ordinarily equivalent to forward-looking utility): suppose that instead of a discounted *sum* of future flow utility, SWB is ordinarily equivalent to a weighted *average* of future flow utility (i.e., discounted sum divided by expected number of remaining years). While under 1B SWB could decline as one has fewer remaining years, under its tweak SWB reflects only the quality of the years that do remain. While both assumptions correspond equally to choice that does not affect the number of remaining years, the implied age-path of SWB could differ dramatically over the life course.

### III.B. Other-regarding preferences

Just as with intertemporal preferences, few SWB applications discuss which other-regarding preference information is captured by the SWB measure. While we are not aware of papers that are explicit and unambiguous, Ludwig et al.'s (2012) discussion of the GSS happiness question suggests that they treat responses as a measure of family well-being: "Another reason we focus on adults is because more is known about measuring SWB of adults than youth ... [SWB] was added to the long-term survey to be one of the key summary measures of the net impacts on families ...." (p. 1507). Easterlin (1995) studies this same GSS happiness question and a variety of international life satisfaction data sources and offers an interpretation that is ambiguous: "Formally, this model corresponds to a model of interdependent preferences in which each individual's utility or subjective well-being varies directly with his or her own income and inversely with the average income of others" (p. 36). Easterlin's mention of "interdependent preferences" sounds like SWB depends on others' *utility*, but the rest of the sentence sounds more like SWB is capturing self-centered utility that depends on relative *income*, as in Frank's (1985) model of status concerns. Similarly, Alesina, Di Tella, and MacCulloch (2004) treat this same GSS happiness question, as well as the Eurobarometer life satisfaction question, as measuring some notion of utility but do not specify if the utility notion is other-regarding or self-regarding (with respondents concerned about how inequality affects their own future prospects): "In this paper, we explore whether and why inequality negatively affects

individual utility even after controlling for individual income. We measure ‘utility’ in terms of survey answers about ‘happiness’” (p. 2010).

Again, as with time horizon, the social-circle interpretation of the SWB question often matters in SWB applications even when it is not discussed. Here are four examples. First, consider the finding mentioned above that giving to others increases short-run happiness (e.g., Dunn, Aknin, and Norton, 2008). How much this can be attributed to altruistic preferences, as opposed to self-signaling (e.g., Bénabou and Tirole, 2006) or warm glow (Andreoni, 1989), depends on the extent to which the SWB data capture other-regarding preference information. Second, a number of papers compare SWB between men and women (e.g., Stevenson and Wolfers, 2009). If SWB is capturing family well-being rather than self-centered utility, and individuals have opposite-sex family members whose well-being affects their own, then the finding that women have lower SWB than men may actually understate the self-centered well-being difference. (That is, if we exclude the relatively higher SWB of the male family member from the female family member’s SWB, then she would report still lower SWB.) Moreover, if men interpret the SWB question as applying to a broader or narrower social circle than women do, and especially if such interpretational differences vary over time, then these interpretational differences confound conclusions from SWB data about self-centered well-being differences between men and women. Third, consider papers that compare SWB between people with and without children (e.g., Deaton and Stone, 2014). Such comparisons are usually construed in terms of self-centered utility; they are much harder to understand if SWB captures family well-being and hence includes a concern for the children.

Finally, consider papers and policy reports that compare mean SWB across countries (e.g., the 2023 *World Happiness Report*; WHR, 2023). When such comparisons are interpreted as national well-being rankings, Assumptions 2–4 underlie the use of mean SWB as a social welfare function. Since the arguments of a social welfare function are almost always self-centered utilities (for a recent review, see Adler, 2019), such applications implicitly assume that SWB is measuring self-centered utility.

#### **IV. Towards Evidence-Based Interpretation of Existing SWB Data**

In the second part of this review we draw on past research, including much of our own work, to give specific, implementable advice to users of SWB data. We start in this section with

evidence-based advice for providers and users of *existing* SWB questions—including time series that have been collected for a while and that, we hope, will continue being collected in the future. In the next two sections we discuss *new* SWB and auxiliary data that we would like to see collected—for example, based on evidence regarding what time horizons and social circles different SWB questions in fact measure.

#### **IV.A. Assessing survey methodology assumptions**

All survey-based research relies on standard survey-methodology assumptions, and SWB research is no exception. That said, careful SWB researchers often open their papers with a note of caution that is specific to SWB data. Deaton (2018), for example, notes in his introduction that in addition to evidence that different SWB measures capture different aspects of well-being and have different correlates (see IV.B below),

The stability and reproducibility of self-reports is an important topic that remains unsettled. For example, the effects of an immediately prior question about the state of the country had a large negative effect on the reports about state of the respondent using the Cantril ladder in Deaton (2011) and Deaton and Stone (2012), and those effects are large enough to compromise its use in at least some contexts, such as time-series monitoring of national wellbeing.

Like other economists, Deaton also refers to a long-standing debate in psychology on the extent to which SWB measures are manipulable by irrelevant cues and, importantly, what one can do about it. Pavot (2018), for example, in a *Handbook of well-being* (written by and for psychologists), discusses the reliability and validity of these measures, and reviews the stable and transient factors that may influence SWB responses. He mentions both earlier research (e.g., Schwarz and Clore, 1983) and subsequent research (e.g., Yap et al., 2017). His own interpretation, more bullish than Deaton’s, is that “these effects are generally minimal, if not nil, and can be further reduced with methodological safeguards.”

The “safeguards” Pavot (2018) and other psychologists recommend include randomizing question order and averaging over SWB measured more than once or using a multi-question battery. Interestingly, psychologists’ specific worry that SWB is disproportionately affected by current mood could, from our (Section II) economics model’s perspective, simply be consistent with certain SWB questions measuring flow utility. We return to this point in Section VI, where we give advice regarding question wording. However, that an immediately prior question has

dramatic effects on SWB responses seems inconsistent with the flow-utility interpretation. It makes interpreting SWB responses as measures of utility (or some stable set of its components) less justified—assuming, of course, that the underlying well-being captured by flow utility is not actually strongly affected by the immediately prior question.

Other survey methodology issues, including nonresponse and self-selection, may also affect the generalizability and interpretability of SWB-based evidence. As Table 1 shows, response rates in some of the most widely used sources of SWB data have fallen to below 50 percent, and sometimes much below. Such low response rates may lead to wrong conclusions from SWB analysis. Heffetz and Rabin (2013) use data from the University of Michigan’s Surveys of Consumers to analyze responses to a yes/no question on happiness during the past week (taken from the HRS; see Table 1) and find that conclusions regarding cross-group differences in reported happiness may depend on whether respondents were difficult or easy to reach. Notably, dividing the sample by number of phone call attempts before respondents completed the survey, they find essentially no gender gap in happiness among the easiest-to-reach respondents, but men are increasingly and significantly happier than women as respondents get harder to reach. Moreover, once covariates are controlled for, the *sign* (in addition to size) of the gap depends on difficulty to reach respondents. As the authors note, this finding suggests that busier women are less happy, both absolutely and relative to men. Thus, if women have become busier over time, this trend might explain Stevenson and Wolfers’s (2009) finding that women have become less happy over time.

As practical advice, Heffetz and Rabin (2013) propose that survey centers could make information about, and paradata from, calling (or other data-gathering) algorithms more easily accessible to researchers. Such data are already collected; researchers should demand them, and use them routinely to assess the robustness of their conclusions.

This and the above survey-methodology advice is straightforward to follow. Building on psychologists’ research, for example, on the effects of “buffer” questions (e.g., Schwarz and Schuman, 1997; but see also Bishop, 1987, for less promising findings), Deaton (2012) shows how thoughtful survey design can at least help assess, and possibly also help adjust for, potential biases. Using the randomization of survey-design elements in the GHWBI, and the addition of “transition” questions before the SWB question, he constructs a “corrected” ladder measure. Of course, as Deaton points out, there is no assumption-free way to make corrections.

As another example, Heffetz and Reeves (2019) use BRFSS life-satisfaction data (see Table 1) and BRFSS call-attempt paradata, both publicly available, to re-examine past results. Revisiting Oswald and Wu (2010)—an important SWB application that we discuss below and that uses the same BRFSS data—they find no evidence that the results are affected by the details of data collection. Similar reexamination of other important SWB results would help both producers and consumers of SWB research focus on the more robust findings in the literature.

#### **IV.B. Interpreting SWB questions as measures of preferences**

The three variants of Assumption 1 (namely, that SWB captures some specific combination of the  $u(c_t)$ 's) represent an early hope in the economics of happiness: Responses to a single SWB question may capture everything that matters for human well-being. Indeed, this hope is implicitly reflected in the title of the paper that launched that literature: Easterlin's (1974) "Does Economic Growth Improve the Human Lot? Some Empirical Evidence." In motivating the use of SWB data to measure "the human lot," Easterlin distinguishes between overall welfare and the narrower concept of economic welfare and notes that economists have assumed that only the latter is objective and measurable. He then goes on to explicitly relate SWB data to welfare: "Happiness corresponds to the broader of these two notions..." (p. 90). Using data from nineteen countries, Easterlin's paper analyzes SWB responses from two types of questions, happiness and life ladder, treating both as direct measures of welfare.

Some influential current researchers (e.g., Frijters, Clark, Krekel, and Layard, 2020) still hold some version of Easterlin's original view that SWB is at least ordinally equivalent to well-being (Assumption 1A, 1B or 1C). Moreover, in their explicit call for SWB-based policymaking, these researchers treat different SWB measures as practically interchangeable, implicitly implying that *different* SWB measures are all ordinally equivalent to well-being.

However, this early idea appears to be less common today. For example, as Stiglitz, Sen, and Fitoussi (2009) put it in their widely publicized, influential report: "Most prominent among these [ambiguities and misconceptions] is the assumption that all dimensions of subjective well-being can somehow be reduced to the single concept of 'happiness'" (pp. 145-146). Emphasizing in their report the three specific dimensions highlighted by the psychologist Ed Diener (1984)—life satisfaction/overall judgement, positive feelings/affect, and negative feelings/affect (see Table 1 for each of these categories of SWB)—they go on to provide examples:

People who experience unpleasant feelings or physical pain could still be very satisfied with their lives if they value the contributions that they feel they are making to society or to some other personal goal. Likewise, the presence of positive affect does not imply the absence of negative affect. Across individuals, the correlation between life satisfaction and positive affect is only around 0.40 ... The correlation among different indicators of negative affect, such as anger and sadness, is also low at the individual level. (p. 146)

The recent availability of large surveys that collect more than a single SWB measure has allowed researchers (both psychologists and economists) to confirm that different SWB questions have different properties both in the cross section—e.g., comparing men versus women (Blanchflower and Bryson, 2023), comparing age groups (Stone, Schwartz, Broderick, and Deaton, 2010), and comparing income groups (Kahneman and Deaton, 2010)—and in the time series (e.g., Deaton, 2012). These papers demonstrate that what data regularities one observes could strongly depend on which SWB measure one examines. Indeed, this point is often the main contribution of a paper, as expressed, for example, in Kahneman and Deaton’s (2010) title—“High income improves evaluation of life but not emotional well-being”—and in Blanchflower and Bryson’s (2023) abstract—“Women say they are less cheerful and calm, more depressed, and lonely, but happier and more satisfied with their lives, than men.”

That different SWB measures give such different pictures implies that they cannot *all* capture overall underlying well-being or, indeed, any *single, well-defined* notion of well-being. Under Assumptions 2–4 and the other survey-methodology assumptions, such evidence implies that at least some SWB measures capture either (i) different variants of Assumption 1 (that is, the  $u(\mathbf{c}_t)$ ’s are aggregated over different time horizons) or (ii) different  $\tilde{u}(\mathbf{c}_t)$ ’s altogether (i.e., different SWB measures are different functions of the components of the  $\mathbf{c}_t$ ’s (as argued by Bernheim, 2016), with different implicit weights on distinct aspects of own well-being or, for example, sometimes including and sometimes excluding aspects of well-being of others besides the survey respondent).

In the next two sections we review evidence that supports both (i) and (ii) and, based on that evidence, we provide advice regarding *new* data collection. We close the present section with a list of recommendations regarding evidence-based interpretation of *existing* SWB questions as preference measures.

**IV.B.1.** *The choice setting: be extra careful when interpreting data about big, important decisions.*



Past research establishes that SWB responses do not cleanly elicit preferences defined as what choices would be under idealized conditions. Benjamin, Heffetz, Kimball and Rees-Jones (BHKR, 2012) present survey respondents with a variety of hypothetical scenarios involving pairwise choices. About each scenario, respondents are asked both which of the two options they think they would choose and which of the two they think would give them higher SWB (for example, a happier life as a whole). Since respondents hold the same beliefs about consequences when answering about choice and when predicting SWB, erroneous beliefs about consequences cannot be responsible for a difference between the two; in that sense, this study design gets closer to idealized (hypothetical) choice conditional on beliefs.

This research finds high concordance between responses to the two types of questions, but also systematic differences, especially in settings that involve what the student respondents perceive as personally important decisions. Such scenarios revolve around studying, working, socializing and sleeping—for example, attending a more fun and social college versus a highly selective one, or following one’s passion versus pursuing a more practical career path. On the other hand, in a placebo scenario involving a choice between two food items that are offered free and for immediate consumption, predicted SWB effectively coincides with predicted choice. That placebo scenario was designed to be similar to small, less important, ubiquitous decisions that involve no consequential differences between the choice options beyond their potential effects on one’s momentary consumption experience.

In a follow-up paper, BHKR (2014), the same authors collect data on an important actual life choice, this time of medical students: how to rank the residencies that they apply to through a centralized clearinghouse that based assignments on these rankings. This choice may again arguably have characteristics close to idealized choice: students deliberate over their decision for months and have a great deal of information and advising available to assist them in becoming well informed. BHKR (2014) find, consistently with the big-decisions finding above, that rankings implied by predicted SWB under different residencies do not coincide with reported choice rankings. Indeed, controlling for predicted SWB under the different residencies, *predicted aspects of the residencies typical SWB questions may neglect*—such as prestige and effect on spousal happiness—help predict reported choice rankings. They find, for example, that relative to the reported choices, anticipated-SWB underweights residency prestige-and-status and

residency desirability for the respondent's significant other, while overweighting anticipated social life and life seeming worthwhile during the residency.

Overall, these studies suggest that using SWB data as a proxy for preferences may be *least* justified in consequential, big-decision contexts—the contexts that, naturally, many economists and (even more so) policymakers are usually most interested in. In such contexts, we recommend extra caution in interpreting findings. On the other hand, for small, personal decisions, SWB may be a better proxy (though not necessarily a *good* proxy, as we discuss below). Such small, ubiquitous decisions are the focus in much applied microeconomics research, as well as in much of psychology and marketing research, for example when studying individual choices in the lab. Indeed, by routinely eliciting predicted SWB and other potential considerations (e.g., doing the right thing) from their choice-making subjects, such researchers could directly *assess* how good SWB is as a preference proxy in their specific context. That practice would also contribute to substantially increasing the currently small body of evidence on the issues we are discussing here.

#### **IV.B.2.** *The choice options: beware of money versus non-money comparisons.*

One important class of beyond-SWB consequences that play a central role in many big decisions relates to money. BHKR (2012) find that in pairwise-choice settings where one alternative involves higher income or more money—compared, for example, to more sleep, a shorter commute, or being around friends—survey respondents are systematically more likely to choose the money alternative than they are to predict it will yield higher SWB. As they point out, this finding suggests that using SWB data to estimate the money valuation of a good (see Section III.A) by comparing the coefficient on the good with that on income in multivariate SWB regressions “may bias these estimates upwards relative to incentivized choice-based estimates” (p. 2107).

This finding has another important implication. While traditional welfare indicators have, justifiably, been blamed for putting disproportionate weight on income and consumption, we have argued (Benjamin, Cooper, Heffetz, and Kimball, 2020) that single-question well-being indicators seem to swing too far in the opposite direction, potentially underweighting income and consumption relative to their weight in preferences. One practical implication is that it is important to collect multidimensional well-being measures—at least until (the possibly never-to-

arrive day when) a comprehensive single-question measure is found. In the absence of such a measure, researchers and policymakers alike should resist the temptation to interpret typical SWB-based valuations as faithfully reflecting preferences, as they seem to underweight dimensions that people may care strongly about.

**IV.B.3.** *The SWB measure: seek comprehensive, and prefer evaluative over affective if limited to a single-question measure.*

In a third paper that examines the SWB-choice relationship, Benjamin, Heffetz, Kimball, and Szembrot (BHKS, 2014) use an alternative method to describe the two hypothetical-choice options: instead of pairwise-choice scenarios, respondents are faced with conjoint tasks, where they choose between “life bundles” that are identical except for explicitly specified aspects (including typical SWB aspects such as happiness and life satisfaction as well as many other aspects of life). The conclusion that emerges from BHKS and both BHKR papers is that evaluative SWB measures are on average closer to choice (hypothetical or real) than affective happiness measures. This evidence-based conclusion is consistent with an intuition that appears to be shared by many researchers (to the best of our knowledge, without direct evidence): Evaluative SWB questions are more comprehensive measures of well-being.

Since the three papers find systematic deviations between choice and anticipated SWB *as measured by widely used SWB questions*, they do not rule out the possibility that anticipated SWB as measured by some as yet unexplored, more comprehensive, evaluative SWB question would reduce, or even eliminate, these choice-SWB deviations. Indeed, their scenario-based and conjoint methodologies could be productively applied to explore new proposals for such questions. In the meantime, as practical advice to users, evaluative-SWB questions do appear to capture preferences (defined by choice) better—but, again, that they do it *better* does not imply that they do it *well*.

**IV.B.4.** *Whose SWB: do not overlook preference for family well-being.*

These three papers also find that measures of family (or couple) well-being are among the strongest predictors of choice—although, of course, they are less relevant in personal consumption decisions with consequences only for self. Such family-SWB measures are not yet widely used in empirical applications; in Section VI below we recommend that they be collected

and studied. BHKS (2014) explores some new measures, and find that “the happiness of your family” and “the overall well-being of you and your family” align with hypothetical choices more closely than currently used measures.

#### **IV.B.5. Summary, and beware of SWB-based pricing!**

Throughout this section, we provide advice as to when SWB may more and less closely align with preferences. We compare big vs. small decisions, money vs. non-money choice options, affective vs. evaluative SWB measures, and measures that focus on self vs. family. Overall, the evidence we review suggests that outside of particular cases, SWB tradeoffs should not be viewed as good proxies for preference-based marginal rates of substitution (MRS)—especially when money is involved, as is the case for pricing exercises.

### **V. SWB and Response Scales**

We now move from advice regarding existing-data interpretation, to advice on future-data collection. We start in this section by discussing theory, evidence, and advice relevant to Assumptions 3 and 4 from Section II.

#### **V.A. Comparing mean SWB across groups: The problem**

We begin with Assumption 4, the assumption that the SWB reports can be averaged (or added up) to obtain a valid measure of a group’s well-being. Comparing means across groups may be justified if a variable is measured on an interval scale. The fundamental problem is that SWB is measured on an ordinal scale (see Table 1).

Using the life satisfaction question on a 0–10 scale in the German Socioeconomic Panel, Ferrer-i-Carbonell and Frijters (2004) regress life satisfaction on sociodemographics using ordinary least squares (OLS), ordered probit, or ordered logit. They conclude that “assuming ordinality or cardinality of happiness scores makes little difference” to the results (quote from their Abstract). As Schröder and Yitzhaki (2017) point out, however, these regressions hold fixed the coding of the response options as 0–10 and therefore do not address the relevant question: do the results depend on which monotonic transformation of the coding is used? Under the assumption that SWB is measured on a continuous scale, Schröder and Yitzhaki provide conditions on the distributions of SWB that ensure that the sign of a univariate regression

coefficient is invariant to any monotonic transformation of the coding. They highlight that these conditions are stringent and rarely hold. For example, the ranking of two groups' mean SWB is invariant under any monotonic transformation if and only if one group's SWB distribution first-order stochastically dominates the other group's.

Bond and Lang (2019) show that the conditions needed to ensure that the signs of regression coefficients are invariant are even more stringent when, instead of being measured on a continuous scale, SWB is measured with a small number of discrete response options, as is typical (see Table 1). In this case, the observed discrete response reflects a latent, continuous SWB level that is categorized into one of the options, but the distribution of latent SWB is unobserved. A researcher can only be sure that the ranking of two groups' mean latent SWB is invariant under any monotonic transformation if one group's latent SWB distribution first-order stochastically dominates the other group's for *every* possible pair of distributions of latent SWB that is consistent with the two groups' observed discrete responses. When no distributional assumptions are made, to infer that group A's mean SWB is higher than group B's requires that *all* responses from group A are higher than *all* responses from group B! Even if distributional assumptions are imposed on latent SWB (e.g., a normal distribution, which underlies ordered probit), two groups cannot be ranked unless their distributions have exactly equal variances. For all practical purposes, these conditions for ranking groups are never satisfied.

Kaiser and Vendrik (2023) argue that Bond and Lang's conclusion is too pessimistic. They point out that reversals in the comparisons of mean SWB between groups, or reversals in the signs of regression coefficients, often require extreme differences between groups in mean latent SWB within a response category. (Bloem, 2021, makes a similar point.) Kaiser and Vendrik argue that it is possible to test for such differences using a continuous SWB measure, since responses on a continuous scale correspond to one particular monotonic transformation of latent SWB. They analyze data from the Dutch LISS panel, in which each respondent answered two happiness questions, one on a 1–10 scale and one on a 1–100 scale, and they treat the latter measure as approximately continuous. Across many sociodemographic groups, Kaiser and Vendrik find only small differences in mean responses on the 1–100 scale within the 1–10 categories.<sup>9</sup> Based on these estimates and other arguments, they contend that sign reversals

---

<sup>9</sup> However, note that a 0–100 response scale may be more similar to a 0–10 response scales than either is to a 3-point or 4-point verbal response scale. If so, Kaiser and Vendrik finding consistent results between a 0–100 and a 0–10

relative to analyses of the continuous scale are likely to be rare. Furthermore, they find that for responses in the 4<sup>th</sup> and higher categories, mean latent SWB increases roughly linearly across response categories. They argue that this finding suggests that the continuous SWB measure may be an interval-scale measure of SWB; if so, then the Schröder and Yitzhaki critique does not apply to analyses of the discrete measure. Despite these more optimistic conclusions, Kaiser and Vendrik emphasize that coefficient *ratios*—which are often central to the conclusions of SWB applications—are often sensitive to even small differences between groups in mean latent SWB within a response category.

### **V.B. Comparing mean SWB across groups: Promising directions**

Three potential solutions have been proposed to the problem of comparing SWB across groups. First, Bloem and Oswald (2022) and Chen, Oparina, Powdthavee, and Srisuma (2022) point out that the quantiles of a distribution are invariant to any monotonic transformation of the variable. They propose econometric approaches to running quantile regressions when the dependent variable is SWB measured with a discrete response scale. A disadvantage of these approaches is that the results of quantile regressions are more difficult to fit into standard cost-benefit approaches to policy analysis.

Second, drawing on evidence that people make choices more quickly when they are farther from indifferent, Liu and Netzer (2023) develop an econometric method to use response times to identify how far latent SWB is from the category thresholds among respondents who choose the same response category. One disadvantage of the method is that it requires response-time data, which is unavailable in standard SWB surveys. On the other hand, such data are already collected for many computerized surveys, or could be collected essentially costlessly. As with other paradata (see Section IV.A), user demand could increase their availability. However, another disadvantage of the method is that it provides little identifying information for intermediate response categories (because being closer to one threshold implies being further from the other threshold). Therefore, somewhat paradoxically, the method is most powerful when applied to SWB questions with only two response options.

---

scale is not very reassuring about consistency between results on a 0–100 (or 0–10) response scale and results on a 3-point or 4-point verbal scale. More evidence is needed on how results with continuous scales compare to 3-point and 4-point verbal scales.

Finally, Benjamin, Cooper, Heffetz, Kimball, and Zhou (2023) propose an approach with two elements. The first element is to elicit SWB on a (quasi-)continuous 0–100 response scale. Doing so eliminates the problem that Bond and Lang (2019) focus on, namely inferring latent SWB from discrete response categories<sup>10</sup>—although it still leaves Schröder and Yitzhaki’s (2017) problem of ordinality. The second element sidesteps the problem of ordinality by giving up on trying to infer the “right” monotonic transformation of SWB. Instead, as we discuss in V.D below, Benjamin et al. develop econometric methods that translate respondents’ SWB to a common interval scale. Importantly—unlike the other approaches discussed above—Benjamin et al.’s (2023) approach adjusts for possible heterogeneity in use of the response scale across respondents (the topic of the rest of this section). A disadvantage of this approach is that it cannot be applied to existing data; it requires data from SWB on a (quasi-)continuous response scale as well as from “calibration questions” that measure respondents’ response-scale use.

### **V.C. Scale-use heterogeneity: The problem**

We now turn to Assumption 3, the assumption that respondents use the response scale in the same way. Oswald (2008) and Kapteyn, Smith, and van Soest (2007, 2009) pointed out that violations of this assumption could threaten many findings of the SWB literature. Following Benjamin, Cooper, Heffetz, Kimball, and Zhou (2023), we distinguish between three types of scale-use heterogeneity across respondents. *Dimensional scale use* refers to how people use response scales for questions about a particular dimension of SWB. Sources of dimensional-scale-use heterogeneity include cultural norms in how much happiness should be expressed and scale re-norming of the sort that may look like hedonic adaptation. *General scale use* refers to how people use response scales for survey questions in general. It may vary due to culture, personality, and context. In between these two, *categorical scale use* refers to how people use response scales for a particular type of question.

---

<sup>10</sup> A longstanding tradition in psychology maintains that 5-point or 7-point scales are best for surveys because respondents cannot reliably discriminate more finely than that. The evidence, however, suggests that while response scales with more than 5 options are similar to each other along several metrics of psychometric validity, respondents rate them as better in terms of allowing them to express their feelings but worse in ease and quickness of use (e.g., Preston and Coleman, 2000). We advocate continuous scales precisely because they maximize how fine-grained respondents can be in expressing their feelings and therefore, theoretically, eliminate the problem of inferring latent SWB from discrete response categories. In practice, however, many respondents likely round their answers. Future research should account for such rounding, as in Giustinelli, Manski, and Molinari (2022).

General scale use differences are called “response styles” in psychology, marketing, and survey research (for reviews, see, e.g., van Vaerenbergh and Thomas, 2013; Weijters, Baumgartner, and Geuens, 2016). The most widely studied response styles are acquiescence and disacquiescence, which refer to giving high and low ratings, and extreme and midpoint responding, which are related to an individual’s variability in ratings. While the response style literature is large, it has received little attention in economics, with a few exceptions. Stone, Schneider, Junghaenel, and Broderick (2019) find that controlling for measures of response style changes the relationships between age and some self-reported SWB measures. Benjamin et al. (2023) find that even responses to questions completely unrelated to SWB, such as judgments of how dark a circle is, are correlated with responses to SWB questions and with demographics, indicating that general-scale-use heterogeneity could confound SWB regressions. Relatedly, Márquez-Padilla and Álvarez (2018), find that countries’ pass/fail grading thresholds in school are related to countries’ mean reported SWB, presumably because both reflect cultural norms of scale use.

The remainder of this subsection discusses dimensional scale use. Bond and Lang (2019) point out that, if researchers are willing to lean heavily on parametric assumptions, it is generally straightforward to test whether two groups use the response scale in the same way using existing data. In their example, latent SWB is assumed to be normally distributed in both groups, and the SWB question has  $N \geq 4$  response categories. The first two response thresholds can be normalized to 0 and 1 in both groups, which pins down both group’s mean and variance. A researcher can then test whether the remaining  $N - 3$  response thresholds are equal. Applying this test to seven SWB questions and group comparisons corresponding to standard findings from the SWB literature, Bond and Lang reject the hypothesis of equal scale-use across groups in all cases.

The main way researchers have tested for dimensional-scale-use heterogeneity is by using “anchoring vignettes”: the respondent is asked to rate the SWB of other people, whose life situations are described in a vignette. King, Murray, Salomon, and Tandon (2004) developed this methodology, laying out the assumptions underlying its use. Kapteyn, Smith, and van Soest (2007, 2009) introduced anchoring vignettes to the SWB literature. We defer discussion of anchoring vignettes to Section V.D.



The assumption of identical dimensional scale use is crucial not only for interpreting SWB regressions, but also within-individual SWB changes over time (Frederick and Loewenstein, 1999). For example, Ubel, Loewenstein, and Jepson (2003) point out that evidence of hedonic adaptation—mean-reversion of SWB after a positive or negative shock—could be partly or fully explained instead by a change in how respondents use the response scale. For example, consider Brickman, Coates and Janoff-Bulmann’s (1978) classic finding that a year after winning, the happiness of lottery winners declines back to the happiness of control participants. This finding could be due to adaptation of happiness or to recalibration of how lottery winners report their happiness.

Lacey et al. (2008) test for such scale-use shifts by comparing ratings made by patients who have a certain disease with ratings made by control participants. They find that patients rated quality of life with their own disease as higher than non-patients did, but patients rated quality of life with other diseases the same as non-patients, suggesting that the rating difference was genuine and not due to changed scale use. Another approach, employed in a number of papers, is to compare previous years’ reported SWB (which are reported with previous years’ scale use) with ratings of previous years’ SWB (which are reported with this year’s scale use). Depending on the paper and context, these papers find mixed evidence on whether scale-use shifts occurred (see Kaiser, 2022, for a literature review and formal development of this approach). This approach generally assumes that respondents have unbiased memories of their situation in previous years, although Fabian (2022) relaxes this assumption by comparing ratings now of previous years’ SWB with respondents’ ratings now of what they *believe* they would have rated their own SWB in previous years—a comparison that holds constant memory of the situation in previous years.

#### **V.D. Scale-use heterogeneity: Promising directions**

To date, the main approach that has been explored in the literature to adjust for scale-use heterogeneity is the anchoring-vignette approach, mentioned in V.C above. This approach relies on two main assumptions. *Vignette equivalence* is the assumption that “the level of the variable [in this context, SWB] represented in any one vignette is perceived by all respondents in the same way and on the same unidimensional scale, apart from random measurement error.” *Response consistency* is the assumption that “each individual uses the response categories for a

particular survey question in the same way when providing a self-assessment as when assessing each of the hypothetical people in the vignettes.” Under these assumptions, survey respondents share a common ranking of the SWB levels in the vignettes (except for measurement error), and respondents’ own SWB can be ranked relative to these levels. In practice, analyses are based on a hierarchical ordered probit regression, in which covariates affect not only the dependent variable but also the categorical response thresholds.

Anchoring vignettes have been used in several SWB applications (Kapteyn, Smith, and van Soest, 2009; Corrado and Weeks, 2010; Angelini et al., 2014; Bertoni, 2015; Montgomery, 2022). For example, using data from the ladder SWB question in Gallup’s World Poll, Montgomery (2022) re-assesses the typical finding that women have higher SWB than men. She finds that women rate the SWB of hypothetical people in vignettes higher than men do, and when she estimates a hierarchical ordered probit regression, she finds that women’s scale-use-corrected SWB is *lower* than men’s.

Even if the key assumptions underlying the anchoring-vignette approach are satisfied, the scale-use-corrected SWB remains an ordinal variable measured in discrete categories, so Bond and Lang’s (2019) concerns still apply. Moreover, the key assumptions are strong, and serious concerns have been raised about them. For example, response consistency would fail if respondents fill in unspecified details about the vignette situation in ways correlated with the respondent’s own situation, and Deaton (2011) points out that vignette equivalence would fail if respondents may differ in their empathy for the individual described in the vignette. Formal tests of the assumptions in the SWB context, based on a variety of different auxiliary assumptions, have generally rejected them (Datta Gupta et al., 2010; d’Uva et al., 2011; Kapteyn, van Soest, and Voňková, 2011; Peracchi and Rossetti, 2013; Montgomery, 2022), although the scale-use adjustment may nonetheless improve the estimates (van Soest et al., 2011).

Recently, Benjamin, Cooper, Heffetz, Kimball, and Zhou (2023) propose another approach to adjusting for scale-use heterogeneity. Unlike the anchoring-vignette approach, which is tailored to SWB measured on a discrete response scale, Benjamin et al.’s approach is tailored to a continuous response scale. As noted above, this feature avoids the concerns of Bond and Lang (2019) (but not Schröder and Yitzhaki’s (2017) concern that SWB is measured on an ordinal scale—a point we return to in the next paragraph). Like anchoring vignettes, Benjamin et al.’s (2023) approach requires data from additional questions to measure scale use, which they

call “calibration questions.” These could be vignettes if the goal is to correct for dimensional scale use, but they could alternatively be questions unrelated to SWB—such as rating the darkness of the circle—if the goal is to correct for general scale use. While correcting for dimensional scale use has the advantage that it corrects for all of the relevant scale-use heterogeneity in an SWB application, the assumptions required for scale-use correction using vignettes are unlikely to hold, as noted above. Moreover, Benjamin et al. find that roughly 2/3 of the overall scale-use heterogeneity in their data is accounted for by general scale-use heterogeneity.

Benjamin et al. propose to conduct empirical analyses by “translating” everyone’s reported SWB to a common scale, namely the mean scale-use in the sample. These translations can be estimated using responses to the calibration questions, and they turn out to be well approximated by linear functions. Benjamin et al. propose assuming linearity, which implies that the common (scale-use-corrected) SWB scale is an interval scale; each respondent’s scale is a shifted and stretched version of the common scale. Using this common scale is a philosophical move: it means that reported SWB on the common scale is the dependent variable of interest, rather than latent SWB, which is some unknown monotonic transformation of it. However, this move has attractive features: it allows the analysis to depend entirely on observables—there is no need to posit an unobservable latent SWB whose existence is arguable—and it eliminates the problems (highlighted by Schröder and Yitzhaki, 2017) of SWB as an ordinal variable. Moreover, ratios of regression coefficients do not depend on whose scale is used (which shifted and stretched version of the common scale). Benjamin et al. develop econometric methods to implement their approach and straightforward tests to explore when and why scale-use adjustment will matter in a particular application.

## **VI. Towards New SWB Measures Guided by Theory and Evidence**

Finally, we build on our past work to recommend approaches for supplemental data collection that go beyond addressing the survey-methodology and scale-use issues from Sections IV and V. We start with advice regarding collecting new single-question SWB measures aimed at eliciting more explicitly focused utility notions. We then discuss our proposed approach for aggregating such single-question measures into a more comprehensive well-being measure.

## VI.A. Theory- and evidence-guided single-question measures

In Benjamin, Debnam Guzman, Fleurbaey, Heffetz, and Kimball (forthcoming), we discuss in some detail two past approaches taken in the literature for evaluating Assumption 1B, i.e., that SWB is ordinally equivalent to  $U_t^{\text{forward}}$ . The first, based on the theory of spatial equilibrium (in equilibrium, across any two geographic locations, there is a marginal resident who is indifferent between staying and moving) and on several additional strong assumptions, uses a variety of datasets and SWB measures to compare mean SWB across U.S. locations (e.g., Glaeser and Redlick, 2009; Oswald and Wu, 2010; Glaeser, Gottlieb, and Ziv, 2016). The second approach, discussed in Section IV.B, compares people’s (hypothetical or actual) choices with what (they report) they believe would maximize their SWB. (We discuss Benjamin, Heffetz, Kimball, and Rees-Jones, 2012, 2014 above; subsequent, closely related work includes Fleurbaey and Schwandt, 2016; Adler, Dolan, and Kavetsos, 2017; Szabó and Ujhelyi, 2017.) Both approaches find systematic evidence against the assumption that SWB is ordinally equivalent to preferences.

Building on previous research (Ross, Eyman, and Kishchuk, 1986; Ralph, Palmer, and Olney, 2011; Junghaenel et al., 2018; Steffel and Oppenheimer, 2009), Benjamin, Debnam Guzman, Fleurbaey, Heffetz, and Kimball (forthcoming) use a third approach. They ask survey respondents a standard or newly authored SWB question, and then ask them what they had in mind—regarding, for example, life domains, social circles, and time horizons—when answering the question (for example: “When answering the [**Happiness**]/[**Well-Being**]/[**Life Satisfaction**] **question**, [which social circles]/[what time periods] did you interpret the question as asking about?”). Consistent with the conclusion from the first two approaches above, respondents’ introspective reports point against interpreting SWB as ordinally equivalent to preferences. For example, reports do not appear consistent with having had in mind the present discounted value of future flow utility (with or without past flows).

Importantly, this third method can shed light on a *broader set of assumptions* than that SWB is ordinally equivalent to preferences. Indeed, respondents’ interpretations of which social circle (e.g., myself versus my family) and time horizon (e.g., right now versus last month) a given SWB question was meant to refer to often differ substantially from researchers’ assumptions. For example, almost none of the standard and new SWB questions examined cleanly captures self-centered forward-looking utility, lifetime utility or flow utility. Moreover,

respondents' interpretations along these dimensions differ both across SWB questions and across sociodemographic groups—with some SWB questions appearing to better capture certain utility notions than others, as well as to reduce cross-group differences in interpretation.

Two important conclusions emerge from this research. First, reported time horizons depend strongly on which SWB question respondents are asked. Second, none of the most widely used SWB questions—life satisfaction, happiness, ladder—shows a pattern consistent with forward-looking or lifetime utility, nor do they exhibit the more flow-like pattern of, for example, a happiness-yesterday question such as the ONS's (see Table 1).

Based on these and other findings in the paper, the authors give practical advice that depends on users' resources and control over their survey data. Users of existing SWB data—and consumers of the resulting research, e.g., policymakers—are cautioned to take into account the ways that these measures seem to deviate from standard utility notions. Those who add their own SWB question to studies are advised to adjust question wording, rather than copying popular wordings—if doing so can bring the SWB notion elicited by the question more in line with the utility notion they want the resulting data to represent. In particular, researchers aiming for a flow utility measure are advised to ask about SWB during a well-defined, recent time interval, as many surveys already do (e.g., the HRS, GHWBI, and ONS; see Table 1), and researchers aiming for a self-centered utility measure are similarly advised to specify the intended social circle. Indeed, the increasingly used question “Overall, how happy did you feel yesterday?” comes the closest to eliciting self-centered flow utility; and the authors' newly phrased question “On a scale from 0 to 10, how would you rate your overall personal well-being?” comes the closest to eliciting self-centered utility (but not flow utility). Finally, well-resourced researchers, officials, or institutions who design, test, and administer their own surveys are advised to use introspective questions, at least in the survey-development stage, to assess how different SWB questions are perceived by different respondents—so that the research and policy recommendations the resulting data eventually yield can be based on assumptions supported by evidence.

## **VI.B. Developing multidimensional utility proxies**

Developing new SWB questions that more accurately elicit specific notions of well-being could go a long way toward justifying research conclusions that rely on specific assumptions.

However, single-question SWB measures, as well-phrased (and, therefore, properly time-integrated and socially weighted) as they may be, may still capture a function of some  $\tilde{u}(c_t)$  other than the  $u(c_t)$  that models preferences in our Section II model. To get at overall well-being, then, researchers may have to measure the individual components of  $c_t$ , then find the right way to weigh and aggregate the different dimensions. In short, they will have to face the thorny problem of well-being multidimensionality.

On this issue, Stiglitz, Sen, and Fitoussi's (2009) write (pp. 144-145):

... quality of life depends on a variety of factors, with none of them claiming absolute priority over the others. The multi-dimensional nature of quality life (as opposed to the scalar nature of income) increases the complexity of the analyses and raises a number of measurement issues. These pertain to whether to use comparable dimensions across people and countries, how to assess the importance of each dimension, which indicators to use to describe accomplishments in the various dimensions, how to present these indicators (e.g. in their raw form, or by standardizing them in various ways), and whether and how to aggregate them.

This quoted summary points to many issues that today, fifteen years later, are still important, open, practical questions that researchers, as well as national and international agencies, struggle with. We now summarize some progress made to date, focusing on our own proposed approach for constructing an *individual-level* multidimensional well-being index.<sup>11</sup>

Our approach was first described in BHKS (2014) and later updated in Benjamin, Cooper, Heffetz, and Kimball (2017), whose presentation we closely follow here. Its point of departure is the traditional framework for measuring aggregate consumption. In that framework, a utility function  $u(c)$  is defined over a consumption vector of  $M$  *market goods*. A traditional aggregate consumption index,  $\sum_{m=1}^M \bar{p}_m c_m$ , weights each good's consumption,  $c_m$ , by its price held fixed at a baseline level,  $\bar{p}_m$ . Driven by small changes in consumption, changes in the index approximate changes in utility (up to a multiplicative constant):  $\sum_{m=1}^M \bar{p}_m \Delta c_m \propto$

$$\sum_{m=1}^M \frac{\partial u(c)}{\partial c_m} \Delta c_m \approx \Delta u.$$

Our approach replaces the vector of  $M$  consumption goods with a more comprehensive vector of  $J$  goods that include everything that matters for preferences. In practice, we identify each of these  $J$  goods with an SWB question about an aspect of well-being. Because the aspects of well-being are not traded in markets (though a subset of the inputs used to produce them may be), price data are unavailable and different individuals may have different marginal rates of

---

<sup>11</sup> For other approaches, see, e.g., Fleurbaey and Maniquet (2011) and Fleurbaey and Blanchet (2013).

substitution (MRSs) across the aspects. Nevertheless, a personal wellbeing index (PWB) can be constructed using each individual's MRSs for the aspects as weights. Specifically, the index is given by  $PWB \equiv \sum_{j=1}^J \frac{\partial u(c)}{\partial c_j} c_j$ , where the marginal utilities (MUs) are defined relative to an arbitrary numeraire aspect. Small changes in this index provide a first-order approximation to changes in the individual's (ordinal) utility (even if the individual's preferences are non-linear).

While BHKS (2014) primarily proposes a long-term agenda, its empirical findings also point to a few readily actionable steps. As discussed above, these findings come from stated-preference conjoint tasks (with the aspects as the attributes) designed to measure the MRSs between different aspects. Perhaps most important, for well-resourced collectors of data—be it governmental agencies or research institutions—these findings suggest prioritizing the measurement of aspects related to family, health, and security; eudaimonia (e.g., having a meaningful life; Ryff 1989) and especially evaluative SWB measures (but see the warning about double counting below); and freedoms and capabilities (which are especially important in the policy context). As concrete examples—but keeping in mind that BHKS's (2014) survey collected data only from a convenience sample of U.S. adults—of the 136 aspects included in their survey, respondents' stated choices rank as the very top aspect a policy one, “freedom from corruption, injustice, and abuse of power in your nation”; while the top-three personal aspects are, in this order, “the overall well-being of you and your family,” “the happiness of your family,” and “your health.”

Indeed, BHKS (2014) propose gathering data on as many aspects as possible, at least initially, even if only for the purpose of eventually selecting a subset of questions for regular inclusion on SWB surveys. For that eventual selection, they propose devoting resources to stated-preference surveys estimating aspects' relative marginal utilities (which summarize the MRSs).

At the same time, as Benjamin, Cooper, Heffetz, and Kimball (2017) emphasize, many challenges remain. They focus on two. First is formulating the set of aspects so that they are comprehensive (i.e., they include all the aspects that matter to individuals), non-overlapping (i.e., no aspect is counted twice, requiring, for example, the identification and elimination of synonymous aspects), and accessible (i.e., respondents can accurately introspect and report about both their own level of an aspect and how it affects their welfare). Second is choosing the response scales—a topic we discussed in detail in Section V. Beyond these two, they briefly

discuss additional challenges that include, among others, respondents' potential discomfort with numbers, scales that may shift over time, aspects' level of generality (which may matter for results), and how to combine objective and subjective aspects. They also point out that existing research suggests that SWB measures may violate accessibility or truthful reporting in systematic ways, for example, due to social-desirability reporting biases (in addition to their sensitivity to contextual details discussed above; see, e.g., Deaton and Stone, 2016). Finally, Benjamin, Cooper, Heffetz, and Kimball (2020) discuss yet additional challenges, such as top-coding.

## **VII. Conclusion**

This review takes pains not only to present the great (indeed, revolutionary) past contribution and future promise of SWB research, but also to catalog some remaining challenges. We hope that rather than becoming disillusioned, readers will join us in viewing these challenges as exciting to-do lists. Indeed, that they are daunting only makes for a greater opportunity for a creative and ambitious researcher.

We close this review with Table 2, which attempts to organize the constructive advice we have given throughout the paper. As economists, we organize the table by two dimensions of cost. First, whether the advice is costless to implement because it relates to interpretation only; or is low-marginal-cost because, although it involves data collection, it tinkers with already existing and already funded survey-design elements; or (potentially very) costly because it requires additional survey time or other additional costs. The second dimension is whether the advice is easy to apply (e.g., it requires implementing reasonably well-understood procedures, such as using an already available survey element or question) or requires developing or adapting relatively new approaches.

=== TABLE 2 AROUND HERE (see next page). ===

These two dimensions also suggest to whom the given advice would be useful. At the top-left quadrant, advice regarding interpretation of already existing data is useful to any producer or consumer of SWB research that uses existing measures. At the bottom left, the



advice about data collection is useful to collectors of data—those engaged in either going-concern data collection or novel data collection—who have different levels of influence over survey design, and can afford different amounts of survey time and other costs. At the bottom right is advice that requires not only potentially additional survey time and budget, but further work developing new instruments. We believe that developing such instruments should be a priority for governments and statistical agencies that want to begin using systematic measurement of SWB as a complement to traditional economic indicators, such as GDP.

Finally, at the extreme bottom right are what we view as the most challenging—and hence, exciting—tasks: developing and implementing calibration questions to help adjust for scale use, introspection questions to help create more targeted single-question measures, and tradeoff questions to elicit weights for a comprehensive well-being index. We hope that researchers will heed our call to help develop these instruments; and that policymakers will lead the charge in including them in large-scale government surveys.

**Table 2: Practical Advice for SWB Data Use & Collection, by Cost and Applicability**

		Readily apply or further develop?	
		Readily apply	Further develop
<b>Interpret or collect data?</b>	<b>Interpret vis-à-vis preferences</b>	Beware of big, important decisions (IV.B) Beware of money vs. non-money (IV.B) Comprehensive > evaluative > affective (IV.B) Do not omit preferences for family WB (IV.B) Beware of SWB-based pricing! (IV.B) Not flow / lifetime / self-centered utility either* (VI.A)	
	<b>Collect: Additional survey time/cost?</b>	<b>No</b>	Randomize question order (IV.A) Use buffer/transition questions (IV.A) Collect paradata; make publicly available (IV.A, V.B)
	<b>Yes</b>	Measure multiple times (IV.A) Measure multiple dimensions (family, health, security; eudaimonic, evaluative; freedoms, capabilities; VI.B)	Use specific language about time horizon / social circles (VI.A) Use (quasi-)continuous scales (V.B) Add calibration questions (V.D) Add introspective questions (VI.A) Elicit tradeoffs/weights (VI.B)

**Notes:** In parentheses, section where advice is discussed.

\*Relates to interpretation as utility notions other than preferences.

## References

- [1.] **Adler, Matthew D.** 2019. *Measuring Social Welfare: An Introduction*. New York: Oxford University Press.
- [2.] **Adler, Matthew D., Paul Dolan, and Georgios Kavetsos.** 2017. “Would you choose to be happy? Tradeoffs between happiness and the other dimensions of life in a large population survey.” *Journal of Economic Behavior & Organization* 139: 60–73.
- [3.] **Adler, Matthew D., and Ole F. Norheim,** eds. 2022. *Prioritarianism in Practice*. Cambridge, UK: Cambridge University Press.
- [4.] **Aghion, Philippe, Ufuk Akcigit, Angus Deaton, and Alexandra Roulet.** 2016. “Creative Destruction and Subjective Well-Being.” *American Economic Review* 106 (12): 3869–3897.
- [5.] **Alesina, Alberto, Rafael Di Tella, and Robert MacCulloch.** 2004. “Inequality and happiness: are Europeans and Americans different?” *Journal of Public Economics* 88 (9–10): 2009–2042.
- [6.] **Andreoni, James.** 1989. “Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence.” *Journal of Political Economy* 97 (6): 1447–1458.
- [7.] **Angelini, Viola, Danilo Cavapozzi, Luca Corazzini, and Omar Paccagnella.** 2014. “Do Danes and Italians Rate Life Satisfaction in the Same Way? Using Vignettes to Correct for Individual-Specific Scale Biases.” *Oxford Bulletin of Economics and Statistics* 76 (5): 643–666.
- [8.] **Angner, Erik.** 2011. “The evolution of eupathics: The historical roots of subjective measures of wellbeing.” *International Journal of Wellbeing* 1 (1): 4–41.
- [9.] **Becker, Gary S., and Luis Rayo.** 2008. “Comment on ‘Economic growth and subjective wellbeing: Reassessing the Easterlin Paradox’ by Betsey Stevenson and Justin Wolfers.” *Brookings Papers on Economic Activity*, Spring: 88–95.
- [10.] **Bénabou, Roland, and Jean Tirole.** 2006. “Incentives and Prosocial Behavior.” *American Economic Review* 96 (5): 1652–1678.
- [11.] **Benjamin, Daniel J., Kristen Cooper, Ori Heffetz, and Miles S. Kimball.** 2017. “Challenges in Constructing a Survey-Based Well-Being Index.” *American Economic Review* 107 (5): 81–85.

- [12.] **Benjamin, Daniel J., Kristen Cooper, Ori Heffetz, and Miles S. Kimball.** 2020. “Self-reported wellbeing indicators are a valuable complement to traditional economic indicators but are not yet ready to compete with them.” *Behavioural Public Policy* 4 (2): 198–209.
- [13.] **Benjamin, Daniel J., Kristen Cooper, Ori Heffetz, Miles S. Kimball, and Jiannan Zhou.** 2023. “Adjusting for Scale-Use Heterogeneity in Self-Reported Well-Being.” Manuscript.
- [14.] **Benjamin, Daniel J., Jakina Debnam Guzman, Marc Fleurbaey, Ori Heffetz, and Miles S. Kimball.** Forthcoming. “What do happiness data mean? Theory and survey evidence.” *Journal of the European Economic Association*.
- [15.] **Benjamin, Daniel J., Ori Heffetz, Miles S. Kimball, and Alex Rees-Jones.** 2012. “What Do You Think Would Make You Happier? What Do You Think You Would Choose?” *American Economic Review* 102 (5): 2083–2110.
- [16.] **Benjamin, Daniel J., Ori Heffetz, Miles S. Kimball, and Alex Rees-Jones.** 2014. “Can Marginal Rates of Substitution Be Inferred From Happiness Data? Evidence from Residency Choices.” *American Economic Review* 104 (11): 3498–3528.
- [17.] **Benjamin, Daniel J., Ori Heffetz, Miles S. Kimball, and Nichole Szembrot.** 2014. “Beyond Happiness and Satisfaction: Toward Well-Being Indices Based on Stated Preference.” *American Economic Review* 104 (9): 2698–2735.
- [18.] **Bentham, Jeremy.** 1789. *An Introduction to the Principles of Morals and Legislation*. London: printed for T. Payne, and Son.
- [19.] **Bernheim, B. Douglas.** 2016. “The Good, the Bad, and the Ugly: A Unified Approach to Behavioral Welfare Economics.” *Journal of Benefit-Cost Analysis* 7 (1): 12–68.
- [20.] **Bertoni, Marco.** 2015. “Hungry today, unhappy tomorrow? Childhood hunger and subjective wellbeing later in life.” *Journal of Health Economics* 40: 40–53.
- [21.] **Beshears, John, James J. Choi, David Laibson, and Brigitte C. Madrian.** 2008. “How are Preferences Revealed?” *Journal of Public Economics* 92 (8-9): 1787–1794.
- [22.] **Bishop, George F.** 1987. “Context Effects on Self-Perceptions of Interest in Government and Public Affairs.” *Social Information Processing and Survey Methodology* 179-199. New York, NY: Springer New York.
- [23.] **Blanchflower, David G.** 2020. “Is happiness U-shaped everywhere? Age and subjective well-being in 145 countries.” *Journal of Population Economics* 34: 575–624.

- [24.] **Blanchflower, David G., and Alex Bryson.** 2023. “The Gender Well-being Gap.” NBER Working Paper No. w31212.
- [25.] **Blanchflower, David G., and Andrew J. Oswald.** 2004. “Well-being over time in Britain and the USA.” *Journal of Public Economics* 88 (7-8): 1359–1386.
- [26.] **Blanchflower, David G., and Andrew J. Oswald.** 2008. “Is well-being U-shaped over the life cycle?” *Social Science & Medicine* 66: 1733–1749.
- [27.] **Bloem, Jeffrey R.** 2021. “How Much Does the Cardinal Treatment of Ordinal Variables Matter? An Empirical Investigation.” *Political Analysis* 30 (2): 197–213.
- [28.] **Bloem, Jeffrey R., and Andrew J. Oswald.** 2022. “The analysis of human feelings: a practical suggestion for a robustness test.” *Review of Income and Wealth* 68 (3): 689–710.
- [29.] **Bond, Timothy N., and Kevin Lang.** 2019. “The Sad Truth About Happiness Scales.” *Journal of Political Economy* 127 (4): 1629–1640.
- [30.] **Brickman, Philip, Dan Coates, and Ronnie Janoff-Bulman.** 1978. “Lottery winners and accident victims: Is happiness relative?” *Journal of Personality and Social Psychology* 36 (8): 917–927.
- [31.] **Chen, Le-Yu, Ekaterina Oparina, Nattavudh Powdthavee, and Sorawoot Srisuma.** 2022. “Robust Ranking of Happiness Outcomes: A Median Regression Perspective.” *Journal of Economic Behavior & Organization* 200: 672–686.
- [32.] **Clark, Andrew.** 2016. “SWB as a Measure of Individual Well-Being” In *The Oxford Handbook of Well-Being and Public Policy*, edited by Matthew D. Adler and Marc Fleurbaey, Chapter 18. UK: Oxford University Press.
- [33.] **Clark, Andrew.** 2018. “Four Decades of the Economics of Happiness: Where Next?” *Review of Income and Wealth* 64 (2), 245–269.
- [34.] **Clark, Andrew, and Andrew J. Oswald.** 2002. “A simple statistical method for measuring how life events affect happiness.” *International Journal of Epidemiology* 31 (6): 1139–1144.
- [35.] **Corrado, Luisa, and Melvyn Weeks.** 2010. “Identification strategies in survey response using vignettes.” <https://doi.org/10.17863/CAM.1107>.
- [36.] **Datta Gupta, Nabanita, Nicolai Kristensen, and Dario Pozzoli.** 2010. “External validation of the use of vignettes in cross-country health studies.” *Economic Modelling* 27 (4): 854–865.

- [37.] **Deaton, Angus.** 2008. “Income, health, and well-being around the world: Evidence from the Gallup World Poll.” *Journal of Economic Perspectives* 22 (2): 53–72.
- [38.] **Deaton, Angus.** 2011. Comment on “Work Disability, Work, and Justification Bias in Europe and the U.S.” In *Explorations in the Economics of Aging*, edited by David A. Wise, 312–314. Chicago, IL: University of Chicago Press.
- [39.] **Deaton, Angus.** 2012. “The Financial Crisis and the Well-being of Americans: 2011 OEP Hicks Lecture.” *Oxford Economic Papers* 64 (1): 1–26.
- [40.] **Deaton, Angus.** 2018. “What do self-reports of wellbeing say about life-cycle theory and policy?” *Journal of Public Economics* 162: 18–25.
- [41.] **Deaton, Angus, Jane Fortson, and Robert Tortora.** 2010. “Life (Evaluation), HIV/AIDS, and Death in Africa.” In *International Differences in Well-Being*, edited by Ed Diener, Daniel Kahneman and John Helliwell, 105–137. UK: Oxford University Press.
- [42.] **Deaton, Angus, and Arthur A. Stone.** 2014. “Evaluative and hedonic wellbeing among those with and without children at home.” *Proceedings of the National Academy of Sciences* 111 (4): 1328–1333.
- [43.] **Deaton, Angus, and Arthur A. Stone.** 2016. “Understanding context effects for a measure of life evaluation: how responses matter.” *Oxford Economic Papers* 68 (4): 861–870.
- [44.] **Decancq, Koen, and Dirk Neumann.** 2016. “Does the Choice of Well-Being Measure Matter Empirically? An Illustration with German Data” In *The Oxford Handbook of Well-Being and Public Policy*, edited by Matthew D. Adler and Marc Fleurbaey, Chapter 19. Oxford University Press.
- [45.] **Di Tella, Rafael, and Robert MacCulloch.** 2006. “Some Uses of Happiness Data in Economics.” *Journal of Economic Perspectives* 20 (1): 25–46.
- [46.] **Diener, Ed.** 1984. “Subjective well-being.” *Psychological Bulletin* 95 (3): 542–575.
- [47.] **Dolan, Paul, and Daniel Kahneman.** 2008. “Interpretations of Utility and Their Implications for the Valuation of Health.” *The Economic Journal* 118 (525): 215–234.
- [48.] **Dunn, Elizabeth W., Lara B. Aknin, and Michael I. Norton.** 2008. “Spending Money on Others Promotes Happiness.” *Science* 319 (5870): 1687–1688.
- [49.] **d’Uva, Teresa Bago, Maarten Lindeboom, Owen O’Donnell, and Eddy Van Doorslaer.** 2011. “Slipping Anchor? Testing the Vignettes Approach to Identification and Correction of Reporting Heterogeneity.” *Journal of Human Resources* 46 (4): 875–906.

- [50.] **Easterlin, Richard A.** 1974. “Does Economic Growth Improve the Human Lot? Some Empirical Evidence.” In *Nations and Households in Economic Growth: Essays in Honor of Moses Abramovitz*, edited by Paul A. David and Melvin W. Reder, 89–125. Cambridge, MA: Academic Press.
- [51.] **Easterlin, Richard A.** 1995. “Will raising the incomes of all increase the happiness of all?” *Journal of Economic Behavior & Organization* 27 (1): 35–47.
- [52.] **Easterlin, Richard, and Kelsey J. O’Connor.** 2022. “The Easterlin Paradox.” *Handbook of Labor, Human Resources and Population Economics*, 1–33. Cham: Springer International Publishing.
- [53.] **Fabian, Mark.** 2022. “Scale Norming Undermines the Use of Life Satisfaction Scale Data for Welfare Analysis.” *Journal of Happiness Studies* 23 (4): 1509–1541.
- [54.] **Falk, Armin, and Thomas Graeber.** 2020. “Delayed Negative Effects of Prosocial Spending on Happiness.” *Proceedings of the National Academy of Sciences* 117 (12): 6463–6468.
- [55.] **Ferrer-i-Carbonell, Ada, and Paul Frijters.** 2004. “How Important is Methodology for the Estimates of the Determinants of Happiness?” *The economic journal* 114 (497): 641–659.
- [56.] **Finkelstein, Amy, Erzo F. P. Luttmer, and Matthew J. Notowidigdo.** 2013. “What Good is Wealth Without Health? The Effect of Health on the Marginal Utility of Consumption.” *Journal of the European Economic Association* 11: 221–58.
- [57.] **Fleurbaey, Marc, and Didier Blanchet.** 2013. *Beyond GDP: Measuring Welfare and Assessing Sustainability*. Oxford: Oxford University Press.
- [58.] **Fleurbaey, Marc, and Francois Maniquet.** 2011. *A Theory of Fairness and Social Welfare*. Cambridge, UK: Cambridge University Press.
- [59.] **Fleurbaey, Marc, and Hannes Schwandt.** 2016. “Do People Seek to Maximize Their Subjective Well-Being – And Fail?” Manuscript.
- [60.] **Frank, Robert H.** 1985. *Choosing the Right Pond: Human Behavior and the Quest for Status*. Oxford University Press.
- [61.] **Frederick, Shane, and Loewenstein, George.** 1999. “Hedonic adaptation.” In *Well-Being: The Foundations of Hedonic Psychology*, edited by Daniel Kahneman, Norbert Schwartz, and Ed Diener. New York: Russell Sage.

- [62.] **Frey, Bruno S., and Alois Stutzer.** 2002. “What Can Economists Learn from Happiness Research?” *Journal of Economic Literature* 40 (2): 402–435.
- [63.] **Frijters, Paul.** 2021. “Measuring Subjective Wellbeing.” In *Handbook of Labor, Human Resources and Population Economics*, edited by Klaus F. Zimmerman, 1–29. Cham: Springer International Publishing.
- [64.] **Frijters, Paul, Andrew E. Clark, Christian Krekel and Richard Layard.** 2020. “A happy choice: wellbeing as the goal of government.” *Behavioural Public Policy* 4 (2): 126–165.
- [65.] **Fujiwara, Daniel, and Paul Dolan.** 2016. “Happiness-Based Policy Analysis.” In *The Oxford Handbook of Well-Being and Public Policy*, edited by Matthew D. Adler and Marc Fleurbaey, Chapter 10. UK: Oxford University Press.
- [66.] **Giustinelli, Pamela, Charles F. Manski, and Francesca Molinari.** 2022. “Tail and center rounding of probabilistic expectations in the health and retirement study.” *Journal of Econometrics* 231 (1): 265–281.
- [67.] **Glaeser, Edward L., Joshua D. Gottlieb, and Oren Ziv.** 2016. “Unhappy Cities.” *Journal of Labor Economics* 34 (2): 129–182.
- [68.] **Glaeser, Edward L., and Charles Redlick.** 2009. “Social Capital and Urban Growth.” *International Regional Science Review* 32 (3): 264–299.
- [69.] **Goodin, Robert E.** 1986. “Laundering Preferences.” *Foundations of Social Choice Theory* 75: 81–86.
- [70.] **Graham, Carol.** 2016. “Subjective Well-Being in Economics.” In *The Oxford Handbook of Well-Being and Public Policy*, edited by Matthew D. Adler and Marc Fleurbaey, Chapter 15. Oxford University Press.
- [71.] **Gruber, Jonathan, and Sendhil Mullainathan.** 2002. “Do Cigarette Taxes Make Smokers Happier?” NBER Working Paper No. 8872.
- [72.] **Gruber, Jonathan, and Sendhil Mullainathan.** 2005. “Do Cigarette Taxes Make Smokers Happier?” *B.E. Journal of Economic Analysis and Policy* 5 (1).
- [73.] **Hausman, Daniel M.** 2012. *Preference, Value, Choice, and Welfare*. UK: Cambridge University Press.

- [74.] **Heffetz, Ori, and Matthew Rabin.** 2013. “Conclusions Regarding Cross-Group Differences in Happiness Depend on Difficulty of Reaching Respondents.” *American Economic Review* 103 (7): 3001–3021.
- [75.] **Heffetz, Ori, and Daniel Reeves.** 2019. “Difficulty of Reaching Respondents and Nonresponse Bias: Evidence from Large Government Surveys.” *Review of Economics and Statistics* 101 (1): 176–191.
- [76.] **Helliwell, John F., and Christopher P. Barrington-Leigh.** 2010. “Viewpoint: Measuring and Understanding Subjective Well-being.” *The Canadian Journal of Economics* 43 (3): 729–753.
- [77.] **Ifcher, John, Homa Zarghamee, and Sandra H. Goff.** 2021. “Happiness in the Lab: What Can Be Learned about Subjective Well-Being from Experiments?” In *Handbook of Labor, Human Resources and Population Economics*, edited by Klaus F. Zimmerman, 1–27. Cham: Springer International Publishing.
- [78.] **Junghaenel, Doerte U., Joan E. Broderick, Stefan Schneider, Marcella May, Alicia Bolton, Kelly P. McCarrier, Larissa M. Stassek, Sarah C. Keithly, and Arthur A. Stone.** 2018. “Frames of Reference in Self-Reports of Health, Well-Being, Fatigue, and Pain: a Qualitative Examination.” *Applied Research in Quality of Life* 13: 585–601.
- [79.] **Kahneman, Daniel, and Angus Deaton.** 2010. “High income improves evaluation of life but not emotional well-being.” *Proceedings of the National Academy of Sciences* 107 (38): 16489–16493.
- [80.] **Kahneman, Daniel, Alan B. Krueger, David A. Schkade, Norbert Schwarz, and Arthur A. Stone.** 2004. “A survey method for characterizing daily life experience: The day reconstruction method.” *Science* 306: 1776–1780.
- [81.] **Kahneman, Daniel, Peter P. Wakker, and Rakesh Sarin.** 1997. “Back to Bentham? Explorations of experienced utility.” *Quarterly Journal of Economics* 112: 357–406.
- [82.] **Kaiser, Caspar.** 2022. “Using memories to assess the intrapersonal comparability of wellbeing reports.” *Journal of Economic Behavior & Organization* 193: 410–442.
- [83.] **Kaiser, Caspar, and Maarten Vendrik.** 2023. “How much can we learn from happiness data?” <https://osf.io/preprints/socarxiv/gzt7a/>.



- [84.] **Kapteyn, Arie, James P. Smith, and Arthur van Soest.** 2007. “Vignettes and Self-Reports of Work Disability in the U.S. and the Netherlands.” *American Economic Review* 97 (1): 461–473.
- [85.] **Kapteyn, Arie, James P. Smith, and Arthur van Soest.** 2009. “Life satisfaction.” IZA Discussion Paper No. 4015.
- [86.] **Kapteyn, Arie, James P. Smith, Arthur Van Soest, and Hana Voňková.** 2011. “Anchoring Vignettes and Response Consistency.” RAND Working Paper Series WR-840.
- [87.] **Kimball, Miles S., Ryan Nunn, and Dan Silverman.** 2015. “Accounting for Adaptation in the Economics of Happiness.” NBER Working Paper No. w21365.
- [88.] **Kimball, Miles S., and Robert Willis.** 2006. “Happiness and Utility.” <http://www-personal.umich.edu/~mkimball/pdf/uhap.pdf>.
- [89.] **King, Gary, Christopher J.L. Murray, Joshua A. Salomon, and Ajay Tandon.** 2004. “Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research.” *The American Political Science Review* 98 (1): 191–207.
- [90.] **Lacey, Heather P., Angela Fagerlin, George Loewenstein, Dylan M. Smith, Jason Riis, and Peter A. Ubel.** 2008. “Are they really that happy? Exploring scale recalibration in estimates of well-being.” *Health Psychology* 27 (6): 669–675.
- [91.] **Lane, Tom.** 2017. “How does happiness relate to economic behavior? A review of the literature.” *Journal of Behavioral and Experimental Economics* 68: 62–78.
- [92.] **Layard, Richard.** 2005. *Happiness: Lessons From A New Science*. London: Penguin Books.
- [93.] **Layard, Richard, and Jan-Emmanuel De Neve.** 2023. *Wellbeing: Science and Policy*. Cambridge, UK: Cambridge University Press.
- [94.] **Lindqvist, Erik, Robert Östling, and David Cesarini.** 2020. “Long-Run Effects of Lottery Wealth on Psychological Well-Being.” *The Review of Economic Studies* 87 (6): 2703–2726.
- [95.] **Liu, Shuo, and Nick Netzer.** 2023. “Happy Times: Measuring Happiness Using Response Times.” University of Zurich Department of Economics Working Paper No. 371.
- [96.] **Loomes, Graham, and Robert Sugden.** 1982. “Regret Theory: An Alternative Theory of Rational Choice Under Uncertainty.” *The Economic Journal* 92 (368): 805–24.

- [97.] **Ludwig, Jens, Greg J. Duncan, Lisa A. Gennetian, Larry F. Katz, Ronald C. Kessler, Jeffrey R. Kling, and Lisa Sanbomatsu.** 2012. “Neighborhood Effects on the Long-Term Well-Being of Low-Income Adults.” *Science* 337: 1505–10.
- [98.] **MacKerron, George.** 2012. “Happiness economics from 35,000 feet.” *Journal of Economic Surveys* 26 (4): 705–735.
- [99.] **Márquez-Padilla, Fernanda, and Jorge Álvarez.** 2018. “Grading happiness: What grading systems tell us about cross-country wellbeing comparisons.” *Economics Bulletin* 38 (2): 1138–1155.
- [100.] **Montgomery, Mallory.** 2022. “Reversing the gender gap in happiness.” *Journal of Economic Behavior & Organization* 196: 65–78.
- [101.] **Nikolova, Milena, and Carol Graham.** 2021. “The economics of happiness.” In *Handbook of Labor, Human Resources and Population Economics*, edited by Klaus F. Zimmerman, 1–33. Cham: Springer International Publishing.
- [102.] **Nussbaum, Martha C.** 2000. *Women and Human Development: The Capabilities Approach*. UK: Cambridge University Press.
- [103.] **Oswald, Andrew J.** 2008. “On the curvature of the reporting function from objective reality to subjective feelings.” *Economics Letters* 100: 369–372.
- [104.] **Oswald, Andrew J., and Stephen Wu.** 2010. “Objective Confirmation of Subjective Measures of Human Well-Being: Evidence from the U.S.A.” *Science* 327 (5965): 576–579.
- [105.] **Pavot, William.** 2018. “The Cornerstone of Research on Subjective Well-Being: Valid Assessment Methodology.” In *Handbook of well-being*, edited by Ed Diener, Shigehiro Oishi, and Louis Tay. Salt Lake City, UT: DEF Publishers.
- [106.] **Peracchi, Franco, and Claudio Rossetti.** 2013. “The Heterogeneous Thresholds Ordered Response Model: Identification and Inference.” *Journal of the Royal Statistical Society Series A: Statistics in Society* 176 (3): 703–722.
- [107.] **Piekalkiewicz, Marcin.** 2017. “Why do economists study happiness?” *The Economic and Labour Relations Review* 28 (3): 361–377.
- [108.] **Preston, Carolyn C., and Andrew M. Colman.** 2000. “Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences.” *Acta Psychologica* 104 (1): 1–15.
- [109.] **Railton, Peter.** 1986. “Moral Realism.” *The Philosophical Review* 95 (2): 163–207.

- [110.] **Ralph, Katherine, Kim Palmer, and Jayne Olney.** 2011. "Subjective Well-being: a qualitative investigation of subjective well-being questions." Working paper for the Technical Advisory Group on 29 March 2012, Office of National Statistics, U.K.
- [111.] **Redelmeier, Donald A., and Daniel Kahneman.** 1996. "Patients' memories of painful medical treatments: real-time and retrospective evaluations of two minimally invasive procedures." *Pain* 66: 3–8.
- [112.] **Ross, Michael, Adele Eyman, and Natalie Kishchuk.** 1986. "Determinants of Subjective Well-Being." In *Relative Deprivation and Social Comparison: The Ontario Symposium, Volume 4*, edited by James M. Olson, C. P. Herman, and Mark P. Zanna. New Jersey: L. Erlbaum Associates.
- [113.] **Ryff, Carol D.** 1989. "Happiness is Everything, or Is It? Explorations on the Meaning of Psychological Well-Being." *Journal of Personality and Social Psychology* 57: 1069–1081.
- [114.] **Schröder, Carsten, and Shlomo Yitzhaki.** 2017. "Revisiting the evidence for cardinal treatment of ordinal variables." *European Economic Review* 92: 337–358.
- [115.] **Schwarz, Norbert, and Gerald L. Clore.** 1983. "Mood, Misattribution, and Judgments of Well-being: Informative and Directive Functions of Affective States." *Journal of Personality and Social Psychology* 45 (3): 513–523.
- [116.] **Schwarz, Norbert, and Howard Schuman.** 1997. "Political knowledge, attribution, and inferred interest in politics: the operation of buffer items." *International Journal of Public Opinion Research* 9 (2): 191–195.
- [117.] **Sen, Amartya.** 1985. *Commodities and Capabilities*. Amsterdam: North-Holland.
- [118.] **Skopec, Laura, Thomas Musco, and Benjamin D. Sommers.** 2014. "A potential new data source for assessing the impacts of health reform: evaluating the Gallup-Healthways Well-Being Index." *Healthcare* 2 (2): 113–120.
- [119.] **Steffel, Mary, and Daniel M. Oppenheimer.** 2009. "Happy by What Standard? The Role of Interpersonal and Intrapersonal Comparisons in Ratings of Happiness." *Social Indicators Research* 92 (1): 69–79.
- [120.] **Stevenson, Betsey, and Justin Wolfers.** 2009. "The Paradox of Declining Female Happiness." *American Economic Journal: Economic Policy* 1(2): 190–225.
- [121.] **Stutzer, Alois, and Bruno S. Frey.** 2010. "Recent Advances in the Economics of Individual Subjective Well-Being." *Social Research*, 77 (2): 679–714.

- [122.] **Stiglitz, Joseph E., Amartya Sen, and Jean-Paul Fitoussi.** 2009. *Report by the commission on the measurement of economic performance and social progress.* <https://ec.europa.eu/eurostat/documents/8131721/8131772/Stiglitz-Sen-Fitoussi-Commission-report.pdf>.
- [123.] **Stone, Arthur A., Stefan Schneider, Doerte U. Junghaenel, and Joan E. Broderick.** 2019. “Response styles confound the age gradient of four health and well-being outcomes.” *Social Science Research* 78: 215–225.
- [124.] **Stone, Arthur, Joseph Schwartz, Joan Broderick, and Angus Deaton.** 2010. “A snapshot of the age distribution of psychological well-being in the United States.” *Proceedings of the National Academy of Sciences* 107 (22): 9985–9990.
- [125.] **Szabó, Andrea, and Gergely Ujhelyi.** 2017. “Choice and Happiness in South Africa.” *Economics Letters* 155: 28–30.
- [126.] **Ubel, Peter A., George Loewenstein, and Christopher Jepson.** 2003. “Whose quality of life? A commentary exploring discrepancies between health state evaluations of patients and the general public.” *Quality of life Research* 12: 599–607.
- [127.] **van Soest, Arthur, Liam Delaney, Colm Harmon, Arie Kapteyn, and James P. Smith.** 2011. “Validating the use of Anchoring Vignettes for the Correction of Response Scale Differences in Subjective Questions.” *Journal of the Royal Statistical Society Series A: Statistics in Society* 174 (3): 575–595.
- [128.] **van Vaerenbergh, Yves, and Troy D. Thomas.** 2013. “Response styles in survey research: A literature review of antecedents, consequences, and remedies.” *International Journal of Public Opinion Research* 25 (2): 195–217.
- [129.] **Weijters, Bert, Hans Baumgartner, and Maggie Geuens.** 2016. “The calibrated sigma method: An efficient remedy for between-group differences in response category use on Likert scales.” *International Journal of Research in Marketing* 33 (4): 944–960.
- [130.] **WHR.** 2023. *The World Happiness Report.* <https://worldhappiness.report>.
- [131.] **Yap, Stevie C. Y., Jessica Wortman, Ivana Anusic, S. Glenn Baker, Laura D. Scherer, M. Brent Donnellan, and Richard E. Lucas.** 2017. “The effect of mood on judgments of subjective well-being: Nine tests of the judgment model.” *Journal of Personality and Social Psychology* 113 (6): 939–961.

## Supplementary Materials

### **A. Behavioral Risk Factor Surveillance System (BRFSS):**

The first reference below has detailed information on response rates and sample information, the second has information on the survey's structure and general history, and the third is a codebook listing the questions asked of respondents. This survey is a repeated cross-sectional survey. Given that the sample sizes in the first reference are disaggregated by survey format, they are not easily interpretable.

**Centers for Disease Control and Prevention.** 2022. *Behavioral Risk Factor Surveillance System 2021 Summary Data Quality Report.*

[https://www.cdc.gov/brfss/annual\\_data/2021/pdf/2021-DQR-508.pdf](https://www.cdc.gov/brfss/annual_data/2021/pdf/2021-DQR-508.pdf).

**Centers for Disease Control and Prevention.** 2023. *About the Behavioral Risk Factor Surveillance System (BRFSS).* [https://www.cdc.gov/brfss/about/about\\_brfss.htm](https://www.cdc.gov/brfss/about/about_brfss.htm).

**Centers for Disease Control and Prevention.** 2012. *Behavioral Risk Factor Surveillance System Questionnaire.* [https://www.cdc.gov/brfss/questionnaires/pdf-ques/2012\\_BRFSS.pdf](https://www.cdc.gov/brfss/questionnaires/pdf-ques/2012_BRFSS.pdf).

### **B. Eurobarometer:**

The first reference below has detailed information on response rates and sample sizes (found by navigating to the “Standard Eurobarometer 99 - Spring 2023 - First results – en” file, under the “Attachments” heading) while the second has information about the survey's history and structure, as well as sampling rules. The third has information on survey questions in Appendix 1. This survey is a repeated cross-sectional survey. Generally, sample sizes for the Eurobarometer have been at least 1,000 per year for countries with more than 1,000,000 inhabitants and around 500 per year for countries with fewer than this number.

**European Union.** 2023. *Standard Eurobarometer 99 - Spring 2023 – Public opinion in the European Union.* <https://europa.eu/eurobarometer/surveys/detail/3052>.

**Postelnicu, Monica.** 2019. “Eurobarometer.” *Encyclopedia Britannica.* <https://www.britannica.com/science/Eurobarometer>.

**Alesina, Alberto, Rafael Di Tella, and Robert MacCulloch.** 2004. “Inequality and happiness: are Europeans and Americans different?” *Journal of Public Economics* 88 (9-10): 2009–2042.

### **C. European Social Survey (ESS):**

The first reference below corresponds to information on response rates and sample sizes, as well as survey questions. The former can be found in the “ESS10 Headline Indicators ed. 2.0” and the latter can be found in the “ESS10 Appendix A7 Codebook ed. 3.1”, both under the “Documents” heading. The second reference has information on the survey's history and structure. This survey is a repeated cross-sectional survey. In the 2020-2022 wave, sample sizes ranged from 875 to 8,725 across countries, with a median of 1,577.

**European Social Survey European Research Infrastructure (ESS ERIC).** 2022. *ESS10 Data Documentation.* Sikt - Norwegian Agency for Shared Services in Education and Research. <https://doi.org/10.21338/NSD-ESS10-2020>.

**European Social Survey European Research Infrastructure (ESS ERIC).** 2020. *About the European Social Survey European Research Infrastructure (ESS ERIC)*.  
<https://www.europeansocialsurvey.org/about/>

#### **D. Gallup-Sharecare Well-being Index (GSWBI):**

The first reference below has information on the survey's structure and history. Email correspondence with the manager of Gallup Analytics has given us information on historical sample sizes across waves. The second reference has information on questions asked in the survey, and the third has very coarse information on response rates. This survey is a repeated cross-sectional survey. Sample sizes ranged from 156,034 to 334,047 with a median of 171,294 until the survey stopped at the end of 2018.

**Gallup News.** 2023. *How Does the Gallup National Health and Well-Being Index Work?*  
<https://news.gallup.com/poll/246200/gallup-national-health-index-work.aspx/>.

**Gallup-Healthways Well-Being Survey.** 2011. *Question List Sorted by Index*.  
<https://news.gallup.com/poll/146822/gallup-healthways-index-questions.aspx>.

**Skopec, Laura, Thomas Musco, and Benjamin D. Sommers.** 2014. "A potential new data source for assessing the impacts of health reform: evaluating the Gallup-Healthways Well-Being Index." *Healthcare* 2 (2): 113-120.

#### **E. Gallup World Poll:**

The reference below has information on response rates to the survey from 2006 to 2016, and lists which countries were included. Email correspondence with the Gallup Analytics manager, described in Section D, has given us information on historical sample sizes. The second reference in Section D provides information on the survey questions. This survey is a repeated cross-sectional survey. Sample sizes have ranged from 103,755 to 229,960 with a median of 148,232 from 2006 to 2022.

**Deaton, Angus.** 2018. "What do self-reports of wellbeing say about life-cycle theory and policy?" *Journal of Public Economics* 162: 18–25.

#### **F. General Social Survey (GSS):**

The reference below has information on current response rates and sample sizes, questions asked, survey structure, and history. This survey is a repeated cross-sectional survey. The sample size of waves has recently been between 2,000 and 5,000, and was 4,145 in 2022.

**Davern, Michael, René Bautista, Jeremy Freese, Pamela Herd, and Stephen L. Morgan.** 2023. "2022 GSS (Cross-section Study) Documentation and Public Use File Codebook." *The General Social Survey (GSS)*.

<https://gss.norc.org/Documents/codebook/GSS%202022%20Codebook.pdf>.

#### **G. Health and Retirement Study (HRS):**

The first reference below has information on historical response rates, sample sizes, and survey structure. The second reference has information on relevant questions asked in the survey, while the third reference further outlines relevant subjective well-being questions asked in the survey. This is a longitudinal panel survey. The sample size has varied from 10,964 to 22,032 since its inception in 1992. The most recently available sample size (in 2020) was 15,723.

**HRS Staff.** 2023. "HRS Core Interview Sample Sizes and Response Rates." *HRS*.  
<https://hrs.isr.umich.edu/documentation/survey-design/response-rates>.

**HRS Staff.** 2022. "Questionnaire on Your Everyday Life and Well-being 2022." *Health and Retirement Study (HRS)*.  
[https://hrs.isr.umich.edu/sites/default/files/meta/2022/core/qnaire/2022\\_SAQ.pdf](https://hrs.isr.umich.edu/sites/default/files/meta/2022/core/qnaire/2022_SAQ.pdf).

**Kimball, Miles S. and Nunn, Ryan and Silverman, Dan.** 2015. "Accounting for Adaptation in the Economics of Happiness." NBER Working Paper No. w21365.

#### **H. Office for National Statistics (ONS) Well-being**

The first reference below has information on the survey structure and questions asked. The second reference provides information on sample sizes. The third reference has response rates for the Labour Force Survey, on which the ONS well-being questions are asked. This is a repeated cross-sectional survey. From 2011 to 2022, the sample sizes for the four well-being questions have varied from 21,310 to 42,800, with a median of 37,405.

**Office for National Statistics (ONS).** 2018. *Personal well-being user guidance*.  
<https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/methodologies/personalwellbeingsurveyuserguide>.

**Office for National Statistics (ONS).** 2023. "UK Measures of National Well-being: July 2023 edition of this dataset."  
<https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/datasets/ukmeasuresofnationalwellbeing>.

**Office for National Statistics (ONS).** 2023. *Labour Force Survey performance and quality monitoring report*.  
<https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/labourforcesurveyperformancequalitymonitoringreport>.