

NBER WORKING PAPER SERIES

WHAT DO HAPPINESS DATA MEAN?
THEORY AND SURVEY EVIDENCE

Daniel J. Benjamin
Jakina Debnam Guzman
Marc Fleurbaey
Ori Heffetz
Miles S. Kimball

Working Paper 28438
<http://www.nber.org/papers/w28438>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2021, Revised April 2023

For helpful feedback, we thank Joan Broderick, Kristen Cooper, Angus Deaton, Dick Easterlin, Danny Kahneman, Andrew Oswald, Matthew Rabin, Alex Rees-Jones, Norbert Schwarz, and Arthur Stone; participants at the Cornell Behavioral Economics Research Group, Cornell Behavioral/Experimental Lab Meetings, Hebrew University Behavioral/Experimental Economics Meetings, AEA annual meetings; BEAM, NBER Summer Institute, and seminar participants at Colorado Boulder, London School of Economics, Louvain la Neuve, Paris School of Economics, Princeton, Stockholm Institute for Future Studies, Warwick, Wharton, UC Berkeley, and USC. We are grateful to Ophir Averbuch, Joel Becker, Samantha Cunningham, Ofer Glicksohn, Arshia Hashemi, Aharon Haver, Yuezhou (Celena) Huo, Mattar Klein, Lev Maresca, Josef McCrum, Ayala Goldfarb, Yotam Peterfreund, Tamar Yerushalmi, and Jianing (Jenny) Ying for excellent research assistance. For financial support, we are grateful to the Samuel Curtis Johnson Graduate School of Management at Cornell University, the School of Public and International Affairs at Princeton University, the National Science Foundation Graduate Research Fellowship Program grant no. DGE-1144153, and NIH/NIA grants R01-AG065364 to Hebrew University, R01-AG040787 to the University of Michigan, R01-AG051903 to UCLA, and P30-AG024928 to Princeton 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 received IRB approval from the relevant institutions and 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 They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2021 by Daniel J. Benjamin, Jakina Debnam Guzman, Marc Fleurbaey, 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.

What Do Happiness Data Mean? Theory and Survey Evidence

Daniel J. Benjamin, Jakina Debnam Guzman, Marc Fleurbaey, Ori Heffetz, and Miles S. Kimball
NBER Working Paper No. 28438

February 2021, Revised April 2023

JEL No. D69,D90,I31

ABSTRACT

What utility notion—e.g., flow/lifetime, self/family-centered—do self-reported well-being (SWB) questions measure? Existing applications make different assumptions regarding the (i) life domains, (ii) time horizons, and (iii) other-regarding preferences captured by SWB data. To obtain relevant evidence, we ask survey respondents what they had in mind regarding (i)–(iii) when answering commonly used—life satisfaction, happiness, ladder—and new SWB questions. We find that respondents’ self-reports differ from researchers’ assumptions, and differ across SWB questions and sociodemographic groups. At the same time, simple SWB-question wording tweaks are effective in moving self-reports towards desired interpretations. We outline actionable suggestions for SWB researchers.

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

Jakina Debnam Guzman
Department of Economics
Amherst College
306A Converse Hall
100 Boltwood Avenue
Amherst, MA 01002
jrd294@cornell.edu

Marc Fleurbaey
Paris School of Economics
48 Bd Jourdan
PARIS 75014
marc.fleurbaey@psemail.eu

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
University of Colorado at Boulder
5776 Glendavon Loop
Dublin, OH 43016
and NBER
miles.kimball@colorado.edu

A Web Appendix is available at

<https://users.nber.org/~heffetz/papers/web-appendix-to-what-do-happiness-data-mean.pdf>

An Individual Responses Web Appendix is available at

<https://users.nber.org/~heffetz/papers/individual-responses-web-appendix-to-what-do-happiness-data-mean.pdf>

1. Introduction

Survey questions about one's well-being, such as questions about happiness and life satisfaction, are increasingly used in empirical work in economics and other social sciences. Some applications of such self-reported well-being (SWB)¹ data assume that SWB measures the utility that would be revealed by well-informed, deliberated choices, were they observed. A small but growing literature casts doubt on this assumption, finding that choices (real and hypothetical) deviate systematically from the option that people believe would maximize SWB (e.g., Benjamin, Heffetz, Kimball, and Rees-Jones, 2012, 2014; Fleurbaey and Schwandt, 2016). However, many applications of SWB data assume that they measure a different notion of utility that captures only part of the preference information relevant for choices, such as flow utility (i.e., the current-period component of utility under the assumption of time-separable preferences) or self-centered utility (i.e., the self-regarding component of utility under the assumption of interpersonally separable preferences). There is little existing evidence to date to assess such assumptions. Many other applications of SWB data make no clear statement about the utility notion being assumed, and yet for many common applications, the conclusions that can be drawn from the empirical analysis hinge on which assumption is made.

In this paper, in order to provide some evidence relevant to a range of assumptions researchers make about SWB data, we conduct two surveys where respondents introspect and report on how they construct their own answers to one or more of nine different SWB questions. Our conclusions are both negative and positive. On the one hand, respondents' reports regarding commonly used SWB questions do not neatly fit what would be predicted by any of the utility

¹ Following Bernheim (2016), we use the term *self-reported well-being* instead of the more standard term *subjective well-being* because Bernheim's term clarifies that we are studying a *measure* of well-being rather than well-being itself, which is inherently unobservable. (However, we retain the familiar abbreviation SWB, which could apply equally to both terms, rather than adopting Bernheim's abbreviation SRWB.)

notions that researchers assume. Moreover, we find heterogeneity across respondents in the extent to which a given utility notion is consistent with their reports. On the other hand, we find that small variations in question wording have large, predictable effects on respondents' reports, which suggests that SWB questions could be designed that come closer to capturing a desired utility notion. We also illustrate how our introspective questions can be used to check for heterogeneity across respondents in how they answer the SWB questions and assess the robustness of conclusions from SWB analyses to such heterogeneity.

Section 2 provides background for our surveys, beginning with a simple theoretical framework. Well-informed, deliberated choices reveal preferences. The objects that enter preferences are one's own and others' lifetime consumption streams in different "life domains." Preferences are additively separable, both intertemporally and interpersonally. We model the ways in which the utility notion captured by SWB data can deviate from these "full" preferences, either because it is limited to special (narrower) cases of this general framework, or because it integrates across domains, time, or others differently than preferences do. We focus on three particular classes of possible deviations: respondents may (i) put weights on various *life domains* when answering an SWB question that differ from the weights that would correspond to their preferences; or they may interpret the SWB question to be asking about (ii) a *time horizon* shorter than their entire life, corresponding to either forward-looking utility (i.e., ignoring past periods) or flow utility (i.e., ignoring all but the current period); or (iii) *social circles* smaller than everyone they may want to take into account in their choices, corresponding to either family-centered or self-centered utility. We briefly review past research on the relationship between SWB and utility, which finds deviations between SWB and choices but does not focus on asking to what extent they are due to (i), (ii), and (iii), and is therefore mostly silent on

whether SWB captures other utility notions. By eliciting respondents' introspections on how they constructed their SWB responses, our surveys can directly test necessary conditions for SWB to capture some of the other utility notions.

In Section 3, we describe the design of our first survey ($N \approx 3,000$), which we conducted among a demographically diverse (albeit not nationally representative) online sample of the U.S. adult population. Our survey begins by asking respondents an SWB question, either one of the three commonly asked questions that have been used in applied economics research—self-reports of happiness, life satisfaction, or where on a ladder of possible lives one would rank—or one of five additional questions that we explore. The SWB question is immediately followed by a sequence of questions about how much weight respondents had put on various domains of life, time periods, and social circles when they answered the SWB question.

In Section 4, we analyze the weights respondents put on different life domains (such as physical health, income and financial security, and family life and relationships), related to (i). These weights also (a) allow us to compare the results from our introspective methodology with past results from revealed- and stated-choice methodologies about the marginal rates of substitution across life domains, and (b) help us calibrate the response scales respondents use when assigning weights in our questions about time horizons and social circles. At the end of the section (4.4), we discuss how we address limitations of the introspective methodology, with reference to examples from earlier in the section. In Sections 5 and 6 we use the questions about time horizons and social circles, respectively, to evaluate how well the SWB measures may correspond to notions of flow versus forward-looking versus lifetime utility (ii), and self-centered versus family-centered versus other-regarding utility (iii).

In each of Sections 4-6, we also address heterogeneity: the extent to which the SWB responses reflect the *same* life-domain, time-horizon, and social-circle weights across sociodemographic groups. Heterogeneity matters because, even if an SWB question were considered to provide an adequate approximation to some utility notion, heterogeneity in how respondents interpret the question generates an additional confound to scientific and policy conclusions researchers often draw.

In Section 7, we describe the design of our second survey ($N \approx 1,500$), which we conducted among a more representative (on age, sex, and ethnicity) U.S. population to test the robustness and replicability of the findings from our first survey. This second survey focused on five SWB questions and was designed to address potential limitations of the first survey. It was also designed to have more statistical power for comparing across SWB questions, which allows us to turn from a descriptive analysis for our first survey to statistical tests of specific hypotheses for our second survey. Comparing findings across the two surveys we conclude that while the design and/or sample differences matter, our main conclusions hold: commonly used SWB questions are not easily interpretable as any of the assumed utility notions, with interpretation differences across both SWB questions and respondent subpopulations; at the same time, slight question-wording changes can dramatically nudge interpretations towards a desired notion.

We report four main empirical findings. The first three, respectively corresponding with (i), (ii), and (iii), concern the relationship between different utility notions and respondents' introspections about how they answered the SWB question. The last addresses heterogeneity. While all four main findings suggest caution when using existing SWB questions as proxies for *any* of the commonly assumed utility notions, each is accompanied by additional findings that also suggest ways to improve existing measures.

First, using the weights on life domains elicited in our first survey, we broadly replicate past findings from the literature, discussed in Section 2, that explores the relationship between choice (real or stated) and SWB. In particular, we find that the weights we elicit are correlated with marginal rates of substitution estimated from stated choices (from Benjamin, Heffetz, Kimball, and Szembrot, 2014), and the magnitude of this correlation is similar to that estimated in past work that elicited life-domain weights underlying SWB responses in a way that did not rely on introspection (Benjamin, Heffetz, Kimball, and Rees-Jones, 2014). These findings suggest that our survey-elicited weights may be reasonable proxies for the life-domain weights underlying SWB responses and are not merely rationalizations of the SWB responses. This and additional findings—e.g., that the weights are greater on domains that an SWB question explicitly asks about—make us more confident in our introspective methodology more generally.

Second, on average across respondents, we find that none of the SWB measures we examine corresponds closely to lifetime or forward-looking utility, and only one—happiness yesterday—closely resembles flow utility. The three commonly used SWB measures that do not explicitly define a time frame—happiness, life satisfaction, and ladder—resemble flow more than lifetime or forward-looking utility, but they seem to also put weight on the past (and to a lesser extent on the future). We also find that the five additional SWB measures we study evoke a variety of distinct time-horizon profiles, some more and some less flow-like than the three commonly used measures. Similar to some of our results about life domains, we show that by rewording SWB questions, researchers can nudge respondents in the direction of a desired utility notion (in this case, a desired time-horizon profile). Indeed, in contrast with a standard happiness question, a time-explicit happiness-yesterday question appears to closely resemble flow utility.

Third, with regard to social circles, on average across respondents the SWB measures we study look more similar to each other than they do with regard to time horizons, with the exception of a new family-well-being question, which differs from the others. Respondents report putting the highest weight on themselves and second highest on their (immediate) family, with wider social circles receiving less or no weight. Thus, our results suggest that common SWB measures capture neither exclusively self-centered nor exclusively family-centered utility but may be consistent with other-regarding utility. We again find that small changes to SWB-question wording can be very effective in causing respondents to adjust the weights in a predictable direction. In particular, replacing “personal” with “family” well-being yields dramatically more family-centered weight profiles.

Fourth, across sociodemographic groups, we find some heterogeneity in the extent to which the SWB questions resemble the above utility notions. For example, they resemble flow utility more for women and the unemployed in our first survey, and more for the young and the low-income in our second survey. Again, question wording matters a great deal: such differences are substantially smaller for the time-explicit happiness-yesterday question than for the standard happiness question, which is vaguer on which time profile it refers to.

Section 8 concludes and discusses broader implications of our findings. While our findings suggest against assuming that common SWB questions are good measures of particular utility notions, they also suggest some readily applicable practical advice for the governmental agencies and researchers that collect SWB data: rather than taking the wording of SWB questions as given, try to tailor them to correspond more closely to the purpose for which they will be used. Among the nine SWB questions we study, two—a newly phrased personal-well-being question, and an existing happiness-yesterday question—come closest to eliciting self-

centered utility, and the latter comes close to also eliciting flow utility; they may be useful points of departure for further refinements. More generally, our findings point toward a research agenda: existing SWB questions—originally designed more than half a century ago for different purposes—should be redesigned to fit their current uses by economists and policymakers as utility proxies.

The Web Appendix provides additional tables and screenshots for both surveys. Web Appendix B discusses and demonstrates how responses to our introspective questions can be used as control variables to assess the robustness of conclusions from a common application of SWB data: a regression of SWB on sociodemographics.

2. Theoretical Framework and Survey Motivation

In this section, we begin by presenting a theoretical framework to clarify different possibilities assumed in the SWB literature regarding what preference information might be captured by an SWB question. Using this framework, we discuss what has been learned in past work on the relationship between SWB and utility and explain our empirical strategy. (Note that the literature also makes other assumptions, such as interpersonal comparability of utility, that we do not discuss here.)

2.1. Preferences

Papers in the SWB literature often assume that, rather than capturing the full preferences that would be revealed by choices under ideal conditions, SWB data capture only components of those preferences. In order for such components to be well defined, we assume that preferences are intertemporally and interpersonally separable: for every period t , an individual's utility is

$$(1) \quad U_t = E_t \sum_{\tau=0}^T \delta_{t,\tau} \sum_{k=0}^K \lambda_k u_k(\mathbf{c}_{\tau,k}),$$

where T is the number of periods in life, K is the number of others whom the individual may care about, $\delta_{t,\tau} \geq 0$ is the discount factor used in period t for weighting the flow utility from period τ , $\lambda_0 \equiv 1$ is the weight on oneself, $\lambda_k \in \mathbb{R}$ is the weight on person $k \neq 0$, and $u_k(\mathbf{c}_{t,k})$ has standard properties and is defined up to a positive affine transformation. The consumption vectors consist of D domains of life: $\mathbf{c}_{t,k} = (c_{t,k,1}, c_{t,k,2}, \dots, c_{t,k,D})'$.²

For convenience, we normalize the discount factors so that $\delta_{t,t} = 1$. In most applications, economists assume exponential discounting ($\delta_{t,\tau} \equiv \delta^{\tau-t}$ for $\tau \geq t$ and $\delta_{t,\tau} \equiv 0$ for $\tau < t$). We allow for more general discount functions to accommodate realistic alternatives, such as present-biased preferences (e.g., Laibson, 1997), and to nest different utility notions as restrictions on the $\delta_{t,\tau}$'s. In economic applications focused on choice behavior, past consumption is irrelevant, so past periods are omitted: $\delta_{t,\tau} = 0$ for $\tau < t$ but $\delta_{t,\tau} > 0$ for $\tau \geq t$. In that case, we refer to U_t as *forward-looking utility*. When past consumption is included—i.e., $\delta_{t,\tau} > 0$ for all τ (for example, $\delta_{t,\tau} = 1$ for all τ)—we refer to U_t as *lifetime utility* (as evaluated given beliefs in period t). In social welfare evaluation, a focus on forward-looking utility is problematic when comparing individuals who differ in age or have different discount functions (e.g., Jackson and Yariv, 2015; Millner and Heal, 2018); for this and other reasons (see, e.g., Adler, 2012, ch. 6), there is a

² Note that for certain aspects of preferences, such as status concerns, it may be observationally equivalent, or nearly so, to model them as domains of life or as other-regarding preferences. We return to this point in footnote 11 in Section 5 when discussing how our survey handles status concerns.

strong tradition in welfare economics of using lifetime utility. Period t 's *flow utility* is $u(\mathbf{c}_t) \equiv \sum_{k=0}^K \lambda_k u_k(\mathbf{c}_{t,k})$ where $\mathbf{c}_t = (\mathbf{c}'_{t,0}, \mathbf{c}'_{t,1}, \dots, \mathbf{c}'_{t,K})'$.

For modeling interpersonal preferences, in equation (1) we made the traditional assumption (as in Edgeworth, 1881) that the individual's utility is a weighted sum of the “internal” utilities of each person, and because we are also modeling intertemporal preferences, we applied this model to flow utilities. This formulation implies that λ_k is also the weight of person k 's lifetime utility in the individual's lifetime utility, and similarly for forward-looking utility (because it assumes that the same discount function is applied to the internal flow utilities of oneself and other people). Many authors have argued that for welfare analysis—for example, for use as inputs into a social welfare function—at least some components of other-regarding preferences should be ignored (e.g., racism); see, e.g., Adler (2013) for a recent review. Indeed, the relevant component of preferences for welfare economics is often considered to be exclusively the self-regarding component (e.g., Hausman, 2012, ch. 8).

2.2. SWB

The preference information elicited by an SWB question depends on how respondents interpret what factors are relevant for answering the question and on how they weight those factors for the purpose of choosing an SWB response. To formalize various possibilities for how SWB data might capture components of preferences, we model an individual's response to an SWB question asked in period t as analogous to the utility function:

$$(2) \quad \tilde{U}_t = E_t \sum_{\tau=0}^T \tilde{\delta}_{t,\tau} \sum_{k=0}^K \tilde{\lambda}_k \tilde{u}_k(\mathbf{c}_{\tau,k}),$$

where we normalize $\tilde{\delta}_{t,t} = 1$ and $\tilde{\lambda}_0 = 1$, but the other SWB discount factors $\tilde{\delta}_{t,\tau}$, the weights on others $\tilde{\lambda}_k$, and the internal flow functions $\tilde{u}_k(\cdot)$ for $k \in \{0,1,2, \dots, K\}$ may or may not be the same as their preference counterparts $\delta_{t,\tau}$, λ_k , and $u_k(\cdot)$. We refer to \tilde{U}_t as the *SWB function*. The resulting value of \tilde{U}_t is then discretized somehow and reported on the SWB question's response scale (see, e.g., Oswald, 2008).

While equation (2) is rather general, applications in the SWB literature typically make assumptions about how $\tilde{u}_k(\cdot)$, $\tilde{\delta}_{t,\tau}$, and $\tilde{\lambda}_k$ relate to $u_k(\cdot)$, $\delta_{t,\tau}$, and λ_k . Note that such assumptions are mutually independent: different assumptions about the three components can co-exist in equation (2). We now formalize common assumptions.

Life Domains. Most applications in the SWB literature appear to assume that the SWB function aggregates consumption across life domains in the same way that the utility function does: $\tilde{u}_k(\cdot)$ is assumed to be a positive affine transformation of $u_k(\cdot)$ (the same transformation for all $k \in \{0,1,2, \dots, K\}$). This assumption is almost never explicit; an exception is Decancq, Fleurbaey, and Schokkaert's (2015) "consistency assumption." Contrary to this assumption, some researchers have argued on theoretical (e.g., Kimball and Willis, 2006; Becker and Rayo, 2008; Bernheim, 2016) or empirical (e.g., Benjamin, Heffetz, Kimball, and Rees-Jones, 2012, 2014; Glaeser, Gottlieb, and Ziv, 2016) grounds that people may weight domains differently when answering an SWB question and when making well-informed, deliberated choices.

Time Horizon. Three different assumptions are commonly made in the literature regarding the time horizon over which preference information is captured by SWB data:

- *Lifetime utility:* $\tilde{\delta}_{t,\tau} = \delta_{t,\tau} > 0$ for all τ .
- *Forward-looking utility:* $\tilde{\delta}_{t,\tau} = \delta_{t,\tau} > 0$ for all $\tau \geq t$ and $\tilde{\delta}_{t,\tau} = 0$ otherwise.

- *Flow utility*: $\tilde{\delta}_{t,\tau} = 0$ for all $\tau \neq t$.

Another possibility would be *backward-looking utility*: $\tilde{\delta}_{t,\tau} = \delta_{t,\tau} > 0$ for all $\tau \leq t$ and $\tilde{\delta}_{t,\tau} = 0$ otherwise. Backward-looking utility might be natural for survey respondents asked to reflect upon their lives. Indeed, some SWB questions elicit SWB over a time horizon that is explicitly backward-looking. Examples include questions about happiness and other emotions yesterday or in the past week (as in many surveys, including the Office for National Statistics Integrated Household Survey and the Gallup-Healthways Well-Being Index) or during a particular episode of the day (as in the Day Reconstruction Method; Kahneman, Krueger, Schkade, Schwarz, and Stone, 2004). Note, however, that the use of such SWB questions is not motivated by some backward-looking utility notion, but rather by their potential to be measures of just-experienced flow utility, or of the integral of flow utility over a recent period of time.³ Because backward-looking utility does not correspond to any standard utility notion, it is not a focus of this paper.

Social Circles. As to which other-regarding-preference information is captured by SWB data, we again view applications in the literature as making one of three assumptions:

- *Other-regarding utility*: $\tilde{\lambda}_k = \lambda_k$ for all k .
- *Family-centered utility*: $\tilde{\lambda}_k = \lambda_k$ for family members k and $\tilde{\lambda}_k = 0$ otherwise.
- *Self-centered utility*: $\tilde{\lambda}_k = 0$ for all $k \neq 0$.

Note that we do *not* require $\lambda_k \neq 0$ when assuming that the SWB function captures other-regarding or family-centered utility. For example, it could be that SWB data capture other-

³ When the flow of hedonic experiences or its integral is measured, the object of measurement is sometimes referred to as “experienced utility” (Kahneman, Wakker, and Sarin, 1997). Advocates of SWB measures often point to a key potential advantage of experienced utility over choice-based measures of preferences, which are referred to as “decision utility,” for welfare analysis: decision utility depends on people’s beliefs about the consequences of their choices (the expectations in equation (2)), but those beliefs may be systematically biased. In contrast, experienced utility may capture the consequences of decision as they are actually experienced, including the effects of learning and adaptation, changes in budget constraints, and changes in beliefs and preferences (e.g., Kahneman, Wakker, and Sarin, 1997; Dolan and Kahneman, 2008).

regarding utility, but the SWB function puts zero weight on person k ($\tilde{\lambda}_k = 0$) because one's preferences put zero weight on that person ($\lambda_k = 0$).

We note that even if some of the assumptions above hold such that the SWB function \tilde{U}_t captures some utility notion U_t , and even if it is the same utility notion across survey respondents, there is a separate question of whether \tilde{U}_t is the same monotonic transformation of U_t across survey respondents. Existing work on scale-use differences in SWB responses includes, e.g., Oswald (2008) and Kapteyn, Smith, and van Soest (2009). In this paper, we focus instead on the distinct question of whether \tilde{U}_t can be interpreted as *any* monotonic transformation of a standard utility notion U_t .

2.3. Prior tests of the assumption that SWB = preferences

Existing literature on whether SWB captures utility focuses on testing whether \tilde{U}_t is a monotonic transformation of U_t , where U_t represents the choices a person would make after deliberation when well informed about the consequences of her options. Work in this literature has taken two approaches.

One approach is 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. Assuming equilibrium and other strong assumptions, if SWB captures preferences, then mean SWB should be equal across locations. Using a variety of datasets with different SWB measures, a number of papers find evidence of non-trivial differences in mean SWB across locations in the U.S. (e.g., Glaeser and Redlick, 2009; Oswald and Wu, 2010; Glaeser, Gottlieb, and Ziv, 2016). Although their paper mainly focuses on changes over time, Glaeser, Gottlieb, and Ziv (2016) point out that this finding is evidence against SWB being a good measure of

preferences. Interestingly, Oswald and Wu (2010) find that mean SWB in a U.S. state is correlated with the state’s “quality of life”: the dollar value of amenities predicted from state-level regressions of wages, rents, and prices on amenities. Glaeser et al. interpret Oswald and Wu’s finding as suggesting that SWB captures the utility benefits from local amenities but not the utility costs of foregone consumption due to lower wages relative to local living costs. In our notation, this hypothesis means that \tilde{U}_t is not a positive monotonic transformation of U_t and, in particular, \tilde{U}_t is more impacted by local amenities, while U_t is more impacted by purchasing power. The difference between them could be driven by differences between any of $\tilde{u}_k(\cdot)$, $\tilde{\delta}_{t,\tau}$, and $\tilde{\lambda}_k$ and, respectively, $u_k(\cdot)$, $\delta_{t,\tau}$, and λ_k .⁴

The other approach is based on survey data on people’s predictions of the SWB consequences of different possible choices. Researchers ask whether people’s (hypothetical or actual) choices coincide with what they believe would maximize their SWB (Benjamin, Heffetz, Kimball, and Rees-Jones, 2012, 2014; Fleurbaey and Schwandt, 2016; Adler, Dolan, and Kavetsos, 2017; Szabó and Ujhelyi, 2017).⁵ For example, Benjamin, Heffetz, Kimball, and Rees-Jones (2014) ask graduating medical students to report their just-submitted ranking over residency programs, and to predict their SWB under each program. While the students usually

⁴ Here are examples: the marginal rate of substitution of consumption of local parks relative to private consumption may be greater in $\tilde{u}_k(\cdot)$ than in $u_k(\cdot)$; the foregone future consumption (from higher living costs) may matter less for SWB than for utility if the SWB discount factors $\tilde{\delta}_{t,\tau}$ weight the future less than the utility discount factors $\delta_{t,\tau}$; and community members’ enjoyment of local amenities may affect SWB more than utility if their weights in the SWB function, $\tilde{\lambda}_k$ for community members k , are greater than their corresponding weights in the utility function, λ_k .

⁵ Note that it is crucial that choices are compared with people’s *beliefs* about what would maximize their SWB, in order to hold relevant features of the situation constant when making the comparison. There is also an interesting but distinct body of research that compares how people make hypothetical choice tradeoffs over outcomes (e.g., health states) with how those outcomes *in fact* affect their SWB (e.g., Dolan and Metcalfe, 2012a). In such comparisons, the differences between (ex ante) choice and (ex post) SWB could be due to changes in information or beliefs between when choice is assessed and when SWB is experienced or due to adaptation (or other changes in preferences or budget constraints) that was not anticipated at the time of choice. Indeed, the differences are often interpreted by advocates of SWB data as evidence that SWB captures welfare-relevant factors that may be missing from choice-based measures of welfare.

choose the option they anticipate would maximize their SWB, there are systematic discrepancies. Other perceived aspects of the residency programs—including stress, the quality of social life, desirability of the location, prestige, and future career prospects—help explain respondents’ choices, controlling for anticipated SWB.

In our reading, the general finding from both approaches is that standard SWB measures capture substantial information about preferences, but do not coincide with the utility that well-informed, deliberated choices aim to maximize. One of the papers, Benjamin, Heffetz, Kimball, and Rees-Jones (2014), specifically aims to rule out, to the extent possible, a $\tilde{\delta}_{t,\tau} \neq \delta_{t,\tau}$ explanation, by carefully controlling the time-horizon interpretation of anticipated-happiness questions.⁶ The happiness-choice discrepancies that remain in the data are thus likely driven by differences between $\tilde{u}_k(\cdot)$ and $u_k(\cdot)$ or between $\tilde{\lambda}_k$ and λ_k .

2.4. Empirical strategy

Choice-based approaches such as those discussed above are most directly useful in testing whether \tilde{U}_t is a monotonic transformation of U_t , but they provide limited information about the components $\tilde{u}_k(\cdot)$, $\tilde{\delta}_{t,\tau}$, and $\tilde{\lambda}_k$ of the SWB function \tilde{U}_t . To study these components directly, our survey asks respondents how they weighted different life domains, time horizons, and social circles when they answered each of several SWB questions.

⁶ Specifically, in addition to asking the students to predict happiness during each residency program, Benjamin, Heffetz, Kimball, and Rees-Jones (2014) elicit predicted happiness during a sequence of explicitly defined future periods (“during the first ten years of your career,” “for the remainder of your career before retirement,” “after retirement”). They then construct from these predictions a best linear predictor of choice, and use it to rank the residency programs. When comparing this ranking with respondents’ just-submitted (choice) rankings, they find slightly smaller discrepancies than when comparing a single happiness question (with a more limited time horizon) with choice, but the discrepancies remain—largely ruling out differences between choice and happiness in the weights assigned to different time horizons as the only reason for the discrepancies.

Because the survey provides no information about the components $u_k(\cdot)$, $\delta_{t,\tau}$, and λ_k of the utility function U_t , it does not allow us to *directly* test $\tilde{\delta}_{t,\tau} = \delta_{t,\tau}$, $\tilde{\lambda}_k = \lambda_k$, or $\tilde{u}_k(\cdot) = a$ positive affine transformation of $u_k(\cdot)$. Thus, we cannot directly draw conclusions about how closely SWB responses correspond to flow utility, self-centered utility, etc. However, we can use our survey to make inferences about whether SWB responses satisfy *necessary conditions* for corresponding to those utility notions.

In particular, our survey is informative about whether $\tilde{\delta}_{t,\tau} \neq 0$ for particular time periods τ and whether $\tilde{\lambda}_k \neq 0$ for particular groups of other people k . Finding $\tilde{\delta}_{t,\tau} \neq 0$ for $\tau < t$, for example, provides evidence against the assumption that SWB captures flow or forward-looking utility, while finding $\tilde{\lambda}_k \neq 0$ for any $k \neq 0$ provides evidence against a self-centered-utility assumption. Our survey also provides some information about how $\tilde{\delta}_{t,\tau}$ varies with τ and how $\tilde{\lambda}_k$ varies with k . Finally, because in our first survey respondents are randomly assigned into one of eight groups, each answering a different SWB question, and in our second survey each respondent answers the same five SWB questions, preferences are either on average the same or in fact the same across SWB-question respondents; any differences in weights found across SWB questions must therefore reflect differences across the SWB functions. This point underlies our falsification tests, which look for sensible changes in weights in response to changes in SWB-question wording. It also means that when we find differences in weights across SWB questions, we can reject the assumption that the SWB questions *all* capture the same utility notion.

3. Survey Design: First Survey

In our first survey, respondents are first asked an SWB question, presented as it would be in a standard survey. Respondents are then faced with a series of follow-up questions that ask them to introspect about how they constructed the response to the SWB question they had just answered.⁷ These follow-up questions appear on subsequent screens, with the original SWB question and answer (e.g., “You answered: 8”) always appearing highlighted at the top of the screen as an easily accessible reminder. The survey ends with standard sociodemographic questions, followed by questions soliciting feedback regarding the survey. See Web Appendix E for screenshots.

In this section, we begin by providing detail on the design of the SWB question that respondents answer. We then discuss the setting of the survey, some general information on our respondents, and their answers to the SWB question. We defer describing the follow-up introspective questions to subsequent sections, where we discuss the design of these questions, their links to the theory from Section 2, and the empirical findings from those questions. (We turn to our second survey in Section 7, where we compare its design and findings to those of the first survey.)

3.1. SWB question

⁷ We are aware of four prior papers that use empirical methodologies similar to ours, asking one of several SWB questions and then asking respondents how they answered it. These papers report a rich set of findings from open-ended questions and interviews (Ross, Eyman, and Kishchuk, 1986; Ralph, Palmer, and Olney, 2011; Jungthaenel et al., 2018) or brief questionnaires (Steffel and Oppenheimer, 1999) that study topics such as the frames of reference respondents use (e.g., comparisons to other people or an earlier time in life) when selecting an answer to an SWB question. Ralph et al. (2011) also study various other aspects of respondents’ reactions to and interpretations of the SWB question. None of this prior work studies the correspondence with utility notions.

After a short “Welcome” screen—where respondents are greeted and asked to take their time, think carefully, and answer each survey question the best they can—each respondent is presented with one of the following eight SWB questions, selected at random:

Ladder:

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.

On which step of the ladder would you say you personally feel you stand at this time?

Please give a number from 0 to 10: _____

Life Satisfaction:

All things considered, how satisfied are you with your life as a whole these days? Please give a number between 0 (extremely dissatisfied) and 10 (extremely satisfied): _____

Happiness:

Taking all things together, how happy would you say you are? Please give a number between 0 (extremely unhappy) and 10 (extremely happy): _____

Family Well-Being:

On a scale from 0 to 10, how would you rate the overall well-being of you and your family? Please give a number between 0 (lowest rating) and 10 (highest rating): _____

Personal Well-Being:

On a scale from 0 to 10, how would you rate your overall personal well-being? Please give a number between 0 (lowest rating) and 10 (highest rating): _____

Meaning & Value:

On a scale from 0 to 10, to what extent do you feel that your life is meaningful and has value? Please give a number between 0 (not meaningful and has no value) and 10 (extremely meaningful and has lots of value):

Options & Possibilities:

On a scale from 0 to 10, to what extent do you feel that your life is full of options and possibilities that you are free to choose from? Please give a number between 0 (extremely limited options to choose from) and 10 (very many options to choose from): _____

Dealing Well:

People's situation in life depends on both the circumstances they have been given and how they deal with these circumstances. To what extent do you feel that you have dealt well so far with the circumstances you have been given in life? Please give a number between 0 ("I have dealt extremely poorly with the circumstances I have been given") and 10 ("I have dealt extremely well with the circumstances I have been given"): _____

The first three questions—Ladder, Life Satisfaction, and Happiness—closely resemble standard SWB questions from large-scale surveys such as the European Social Survey, the General Social Survey, the Gallup World Poll, and the U.K.'s Office for National Statistics Integrated Household Survey.⁸ The Ladder and Life Satisfaction questions are considered all-purpose evaluative measures. While happiness could be primarily an emotional state, the specific "Taking all things together" Happiness question above also likely has an evaluative component. The three questions, or close variations on them, have been widely used by economists.

Of the remaining five questions, four are new questions that, to the best of our knowledge, have not been previously used in applied work; and one (Meaning & Value) has variants that, as we discuss below, are the focus of a literature on "eudaimonic" SWB. We include them in the survey with the general aim of exploring their potential to "do better" for our purposes than commonly used questions. By "do better" we mean that they may (1) more closely

⁸ Our Life Satisfaction question exactly matches the wording from the Integrated Household Survey's (IHS) life satisfaction question, one of the IHS's four SWB questions. In our second study, we include an additional happiness question that matches the wording from the IHS's happiness question.

track a clear utility notion or (2) be interpreted more comparably across respondent groups. SWB questions meeting these criteria would make it easier to interpret applied work.

The fourth question—Family Well-Being—has been chosen in light of evidence of its potential to satisfy criterion (1). Benjamin, Heffetz, Kimball, and Szembrot (2014) find that a version of this question does best as a predictor of hypothetical choice among 113 questions they study, in a survey design and regressions that attempt to control for all other questions. It may therefore correspond most closely to stated preferences.

The fifth question—Personal Well-Being—is a version of Family Well-Being that takes “family” out of the picture, replacing it with “personal.” Like Family Well-Being, it is included, first, because of its potential to satisfy criterion (1): a measure that uses the phrase “personal well-being” may better capture a more self-centered utility notion, exclusive of any other-regarding preferences (even towards immediate family). Second, we include it in order to explore to what extent an explicit reference to “family” versus “personal” well-being affects how respondents construct their answer.

The sixth and seventh questions—Meaning & Value, and Options & Possibilities—are included for three purposes. First, we explore whether they better satisfy criterion (2) above: Are they interpreted more similarly across respondents than standard SWB questions? Second, related to criterion (1) above, there is a literature arguing that such eudaimonic SWB questions should be included in standard well-being surveys because the dimensions of well-being they assess are not fully captured in standard evaluative SWB questions (e.g., Ryff, 1989; White and Dolan, 2009). Indeed, on these grounds (Dolan and Metcalfe, 2012b), the Office for National Statistics Integrated Household Survey includes a question “Overall, how worthwhile are the things that you do in your life?” Moreover, there have been proposals to include eudaimonic

questions in a multiple-question SWB index that may more closely capture preferences (e.g., Benjamin, Heffetz, Kimball, and Szembrot, 2014). In order for such an index to correspond to a clear utility notion, each question in the index would have to be interpreted similarly in terms of time horizon and social circle—a precondition we can test by including the questions in our survey. Third, these questions serve as falsification-test questions: unlike the first five SWB questions, these two ask about specific domains of life; since a follow-up question asks about the weights a respondent gave to domains that include these, these questions allow us to investigate whether respondents have attentively read and understood the SWB question.

Finally, the eighth question—Dealing Well—attempts to capture the difference between respondents’ evaluation of their *situation*, and their evaluation of the way they have *dealt* with the (exogenous) circumstances life threw at them. Standard evaluative SWB questions, including versions of the Ladder, Life Satisfaction, and Happiness questions above, are typically understood as evaluating an individual’s situation. The switch to evaluating how an individual has responded to circumstances may help satisfy criterion (2), by focusing on something that may be more comparable across individuals who face different circumstances and by specifying the question’s time horizon: the past. At the same time, it may interfere with attempting to satisfy criterion (1), because it is not likely to elicit a (comprehensive) utility notion.

3.2. Survey setup and respondents

The survey was conducted during June 13–30, 2014. Our respondents were recruited by Clear Voice Research, a private firm that invites individuals to “get paid to take surveys and share your opinions about the products and services you use every day” (see <http://www.clearvoicesurveys.com>). To complete the survey, respondents were required to

answer the SWB question (on the second survey screen) and to go through the rest of the screens, although they were allowed to skip all subsequent questions. 3,926 respondents started our survey, and 3,040 completed it, resulting in between 359 and 397 complete responses for each of the main eight SWB questions. We aimed at a sample that, while not a random sample, resembles the adult (18+) U.S. population on basic sociodemographic characteristics. Web Appendix A compares our 3,040 respondents with the U.S. population as described by the U.S. Census and other official sources. While our respondents roughly match the population on sex and marital status, they are more educated and middle-aged, with household income that is more concentrated in the \$40,000–\$80,000 range, more Northeast and less South, more White, with somewhat larger households, and with higher participation in the labor force. Median survey completion time was 14 minutes.

3.3. Responses to SWB question

Figure 1 reports histograms summarizing responses to the main SWB questions, by SWB question and pooled. The median response is 8 on a 0–10 scale in all but the Ladder (median = 6) and the Meaning & Value (9) questions (5th and 95th percentiles are 1–4 and 10 in all questions).⁹ The median time to answer the SWB question was 12.6 seconds (5th and 95th percentiles were 5.4 and 57.6 seconds).¹⁰

⁹ To the extent that top-coding is a worry, Ladder has an advantage over other questions, with the lowest share of respondents reporting 10. On the other hand, these results suggest that the Meaning & Value question—a question whose specific wording we authored (see previous section)—should perhaps have been phrased, if possible, in a way that would push responses away from 10.

¹⁰ Median time to answer each of the eight SWB questions ranged from 10.1 seconds (henceforth, s) to 19.7s, and the variation is almost entirely explained by question length: a regression (with eight observations) of median response time on number of words (or letters) in each SWB question yields an estimated median response time = $5.5s + 0.18s$ per word (or $5.5s + 0.04s$ per letter), with $R^2 > 0.96$. We are not sure what to conclude from these relatively quick responses to complex questions. It is consistent with respondents answering in accordance with a heuristic (such as relying on current feelings; Schwarz and Strack, 1999), but it is also consistent with respondents

4. Weights on Life Domains and Tests of Introspective Methodology

We begin by analyzing reports of the importance, or weight, respondents thought different life domains had on their SWB answer. (Following past research, our survey refers to life domains as “aspects of [people’s] life / situation.”) The results are useful in assessing our introspective method, both by comparing them across different SWB questions and by comparing them with aspect-weight findings from past research. We also use the range of numbers assigned as weights to calibrate what is a relatively “low” and “high” weight in responses to other introspective questions we study in subsequent sections.

Specifically, we examine our survey respondents’ answers to the following question:

People often attribute unequal importance to various aspects of their life. When answering the [Life Satisfaction] Question, how much weight do you think the following aspects of your situation had on your answer?

Here and in other parts of the survey, “[Life Satisfaction]” was replaced by one of the other seven SWB-question titles when relevant. Other elements of the survey screen, including the highlighted SWB question and answer, were held fixed throughout the survey. The question was followed by fifteen domains of life¹¹ in random order, and a sixteenth “Other (please specify)”

already having a rough sense of the answer to the question before being asked. Relatedly, we also find that none of the SWB questions is judged difficult to answer. The first survey question after the main SWB question was: “How difficult was it to answer the [Life Satisfaction] Question?” (with “[Life Satisfaction]” replaced with the title of the SWB-question version that each respondent answered). Overall median and mean response were 11 and 28.0 on a 0–100 scale; by SWB question, median and mean response were in the ranges 8–21 and 24.3–31.5, respectively, with the Happiness and Ladder (and perhaps also the Life Satisfaction) questions being rated on average as slightly easier to answer than the Dealing Well (and perhaps also the Family Well-Being) question.

¹¹ We include “Social status” among the domains of life, even though, as noted in footnote 2, status concerns could alternatively be modeled as other-regarding preferences. The reason we do so is that we do not think respondents’

entry (always at the bottom), each with a slider labeled from “Not at All” to “A lot.”¹² The domains we asked about can be mapped onto those that past work in the SWB literature has identified as important (e.g., Van Praag, Frijters, and Ferrer-i-Carbonell, 2003; Heller, Watson, and Iles, 2004; Easterlin, 2006; Kapteyn, Smith, and van Soest, 2009). Due to relatively few responses, here and in the rest of the paper we do not include the “Other” option in the analysis.

4.1. Mean responses by SWB question: general patterns

Figure 2 Panel A shows the average weights assigned by respondents to the domains (calculated by dividing the unnumbered slider scales into 101 equidistant points), ordered from highest to lowest, separately across the eight SWB questions (leftmost graph) or smaller subsets of questions (middle and rightmost graphs). Each point estimate is based on roughly the same number of observations (359–397), resulting in similar standard errors (the capped bars).

We note three general observations regarding the leftmost graph (“All SWB questions”). First, the means vary widely across domains and SWB questions, from around 35 to around 75.

Second, across the eight SWB questions the vectors of mean domain weights are highly correlated, with correlations ranging from 0.89 to 0.99 (in the next subsection we discuss outliers, such as the “Purpose & meaning” domain in the Meaning & Value question). Our respondents thus report, on average, a similar domain weighting scheme across a wide range of SWB questions. The domains “Income & financial security,” “Family life & relationships,”

concerns about social status are likely to be reflected in their responses to our social-circles questions (described in Section 6 below). Note, however, that in order to accommodate concerns about relative consumption, our model would need to be extended to allow the internal flow utility for oneself, u_0 , to depend on $c_{t,0} - c_{t,k}$. We do not pursue this extension because the exposition of the model in Section 2 is clearer without it.

¹² We considered, but ultimately decided against, assigning numerical values to the slider locations and constraining the sum of the numbers across sliders to be 100. We decided to use the “Not at All” to “A lot” scale because we believed respondents would find it more intuitive and thus be able to introspect more accurately. (In our second survey, respondents do distribute 100 points among categories, with extensive instructions and tests for comprehension. See Section 7 below.)

“Physical health,” “Mental health & emotional life,” and “Security regarding life & the future”—in this order—dominate the top of the figure.

Finally, the domains’ relative weights appear to broadly replicate conclusions from the literature discussed in Section 2.3 above. Exploring the “SWB = preferences” hypothesis, that literature generally finds that standard SWB measures are closely related, but are not identical, to preferences. For example, looking at nine domains related to medical residencies (e.g., prestige), the Benjamin, Heffetz, Kimball, and Rees-Jones (2014) study discussed in Section 2.3 found correlations of 0.69–0.85 (depending on the SWB question) between anticipated-SWB-based and choice-based MRS estimates (where the MRSs are relative to the average domain); using our notation from section 2, those correlations are between \tilde{U}_t MRSs and U_t MRSs.¹³ To investigate the information captured by our slider-based domain weights, we similarly compare them with MRS estimates for U_t from Benjamin, Heffetz, Kimball, and Szembrot (2014, henceforth BHKS), who used a hypothetical-choice survey to estimate the MRSs of 113 aspects of life.

While the 15 domains in our survey do not all perfectly match aspects on BHKS’s list, 12 have

¹³ Formalizing the MRSs for U_t and \tilde{U}_t in terms of our theoretical framework requires some additional assumptions. The simplest such assumptions would be: when a respondent is asked about a change in a life domain, the respondent imagines the change (i) occurs only in the current period t , and (ii) only affects the respondent herself. Assumptions (i) and (ii) would allow us to ignore possible differences across U_t and \tilde{U}_t on the time-horizon and social-circle dimensions when analyzing MRSs across life domains. Specifically, under (i) and (ii), the relevant MRSs for U_t and \tilde{U}_t when analyzing a small change in domain d relative to domain d' are simply the corresponding MRSs for oneself, u_0 and \tilde{u}_0 , respectively:

$$\frac{\partial U_t}{\partial c_{t,0,d}} \bigg/ \frac{\partial U_t}{\partial c_{t,0,d'}} = \frac{\partial u_0}{\partial c_{t,0,d}} \bigg/ \frac{\partial u_0}{\partial c_{t,0,d'}} \quad \text{and} \quad \frac{\partial \tilde{U}_t}{\partial c_{t,0,d}} \bigg/ \frac{\partial \tilde{U}_t}{\partial c_{t,0,d'}} = \frac{\partial \tilde{u}_0}{\partial c_{t,0,d}} \bigg/ \frac{\partial \tilde{u}_0}{\partial c_{t,0,d'}}.$$

We think assumption (i) is reasonable given the wording used in prior surveys (e.g., Benjamin, Heffetz, Kimball, and Szembrot (2014) ask respondents to imagine a change “over the next four years”). Assumption (ii), however, is likely violated for most of the domains we study. For example, a change in a respondent’s domains “income and financial security” or “family life and relationships” entails a change in these domains for her family too. For such a domain, its MRS with respect to another domain could differ across U_t and \tilde{U}_t even if $u_k = \tilde{u}_k$ for all k , due to the weight on family members in the SWB function (the $\tilde{\lambda}_k$ ’s for family members) differing from their weight in the utility function (the corresponding λ_k ’s). More generally—if neither assumption (i) nor (ii) holds—as we emphasize in Section 2.4 above, the MRSs across life domains for U_t and \tilde{U}_t could differ due to differences in any of the components $u_k(\cdot)$, $\delta_{t,\tau}$, or λ_k , relative to $\tilde{u}_k(\cdot)$, $\tilde{\delta}_{t,\tau}$, or $\tilde{\lambda}_k$, respectively.

reasonably similar BHKS counterparts. For these, the correlation between the weights in our survey (averaged across all SWB questions) and BHKS’s MRS estimates is 0.77 (the rank correlation is 0.80)—well within the above range and remarkably high, given that the two studies have entirely different designs.¹⁴

We conclude that on average, the domains’ relative weights from our survey are as related to existing estimates of the domains’ MRSs for U_t as past estimates of their MRSs for \tilde{U}_t , estimated using a different methodology. In the rest of this paper, we therefore proceed under the working assumption that our slider-based domain weights capture substantial information about the MRSs for \tilde{U}_t , and that similarly, our slider-based time-horizon and social-circles weights provide substantial information about $\tilde{\delta}_{t,\tau}$ and $\tilde{\lambda}_k$.

4.2. Comparing across SWB questions

The outliers within the high correlations across the eight SWB questions are best seen in Panel A’s rightmost graph. They suggest that we pass the falsification test outlined in Section 3.1. Specifically, three of the clearest visual outliers suggest that respondents react to the wording of both the domains and the SWB questions as one would expect from attentive respondents: the domains “Purpose & meaning” and “Live personal values” get unusually high weights in the Meaning & Value question; and the domain “Possibilities in life” gets an unusually high weight in the Options & Possibilities question. (The Meaning & Value question

¹⁴ BHKS’s closest 12 “private-good” aspects, in an order corresponding to the domains in Figure 2 Panel A, are: Your financial security (relative marginal utility estimate = 0.34); The quality of your family relationships (0.37); Your health (0.42); Your mental health and emotional stability (0.34); Your sense of security about life and the future in general (0.33); Your sense that your life is meaningful and has value (0.32); You being a good, moral person and living according to your personal values (0.40); You having many options and possibilities in your life and the freedom to choose among them (0.32); Your physical safety and security (0.28); The overall quality of your experience at work (0.10); Your social status (−0.06); Your sense that you are making a difference, actively contributing to the well-being of other people, and making the world a better place (0.29).

stands out in lying to the right of the rest of the pack not only on these two domains but also on others that could reasonably be thought of as related to meaning and value, such as “Volunteering, activism” and “Family life & relationships.”)

Finally, the middle graph highlights the three traditional SWB questions. They effectively coincide on almost all domains, suggesting that overall, respondents assign similar weights across these questions. The few exceptions seem largely consistent with the view that the Life Satisfaction and Ladder questions capture a less emotional notion of well-being than Happiness. For example, for the Ladder and Life Satisfaction questions, respondents give higher weight to “Income and financial security” than “Mental health & emotional life,” while for the Happiness question, respondents give them essentially identical weights.

4.3. Comparing across respondent sociodemographics

Figure 3 Panel A is based on the same data as Figure 2 Panel A, but responses are pooled across all eight SWB questions¹⁵ and are then split by respondents’ age (three groups), sex (two), income (three), and employment status (two, for labor-force participants only). We focus on these four sociodemographic dimensions because they have received much attention in the SWB literature.

It is important to remember that unlike in Figure 2, where the SWB-question-specific curves are based on respondents who are randomly assigned into one of the eight SWB

¹⁵ Web Appendix D reproduces Figure 3 three times, for three disjoint subsets of the eight questions: (a) Ladder, Life Satisfaction, and Happiness; (b) Personal and Family Well-Being; and (c) Meaning & Value, Options & Possibilities, and Dealing Well. While standard errors are wider than in Figure 3, the appendix figures suggest that the new SWB questions in subsets (b) and (c) do not differ in sociodemographic heterogeneity of weights from the standard SWB questions in subset (a)—see our criterion (2) in Section 3.1 above. The similarity of patterns across subsets (a)–(c) motivates our decision to pool the SWB questions when comparing across sociodemographic groups here and in subsequent sections. We return to this point in Section 7, where we discuss our second survey, which has more power to detect cross-question cross-sociodemographic differences.

questions, in Figure 3 assignment into sociodemographic groups is likely to be correlated with other observable and unobservable characteristics of the respondents. As a result, the groups may systematically differ, for example, in how they use the slider response scales. When interpreting the figure—and all other sociodemographics-based comparisons in the rest of this paper—we therefore focus on cross-group differences that could not be explained by biases that could be characterized as merely stretching and shifting the response scale (in the same way across question items). We instead focus on differences between groups in the ordinal ranking of items.

We begin, in Figure 3 Panel A, by noting the overall (ordinal) similarity across sociodemographic groups: while some groups systematically use a wider range of the 0–100 scale than others, in all four graphs the relative ranking of domains is generally maintained across the groups. This too appears consistent with BHKS’s finding of limited cross-group variation in relative marginal-utility rankings. The exceptions, however, again suggest that respondents respond meaningfully to our introspective survey. “Physical health” is the most important domain for those above 55, while for the rest, “Income and financial security” and “Family life and family relationships” are both more important. Women report significantly less weight on “Work and relationships with co-workers” than on “Quality of the environment,” while men report essentially the same weights on both. Most dramatically, “Work and relationships with co-workers” drops in reported weight among unemployed respondents relative to employed ones.

4.4. Introspective methodology

Having illustrated our survey methodology in the context of life domains, we now discuss two limitations, and our approaches to dealing with them.

First, our data are respondents' reported introspections regarding the response they have just given to an SWB question. One potential concern is that respondents are rationalizing rather than introspecting accurately. The consistency between the weights we estimate for life domains with those from related past results obtained with different methodologies (in 4.1 above) provides some reassurance that the introspections are informative about the considerations underlying the SWB response. Nonetheless, reported introspection may miss influences on SWB responses that respondents are unaware of or are unwilling to truthfully report. We highlight, however, that using SWB data in the first place relies on the assumption that people can introspect accurately and do report truthfully about their internal state. Indeed, the considerations that led to one's SWB response—which are what we aim to measure with our introspective questions—are arguably *more* cognitively accessible than the overall evaluation of one's situation on a 0–10 scale required for generating the SWB response. Finally, to verify attentiveness and understanding, we conduct various falsification tests (as in 4.2 above, and throughout the paper), and we embed attention checks in our second survey (see Section 7).

Second, while we can ordinally compare the weights respondents put on various considerations, we need to be careful when drawing conclusions regarding their magnitudes. Although we anchor the 0–100 response scale for the self-reported weights by labeling 0 as “Not at all” and 100 as “A lot,” there is no clear cardinal interpretation of the scale, and response noise would drive mean weights away from the extremes even if many respondents truly assigned weights of 0 or 100. We do sometimes draw inferences that magnitudes are non-zero but only when we see that respondents' mean weights are substantially larger than mean weights on other introspective questions. For example, we conclude in the next section that none of the SWB measures has a time-horizon profile corresponding to flow utility because the weights on time

periods other than the present are all larger than the scale midpoint of 50 and therefore larger than around one-third of the weights on life domains in this section. We also make comparisons that, instead of a cardinal interpretation of the weights, rely on weaker assumptions. For example, we compare mean weights across (randomly assigned) SWB questions, which only requires that the SWB question does not affect respondents' use of the scale for answering the introspective questions. As another example, we compare ordinal rankings of the mean weights assigned by different groups of respondents (which suggests, but does not straightforwardly translate to, an ordinal ranking at the individual level). Finally, in our second survey (Section 7), we use an alternative design where weight magnitudes are easier to interpret.

The structure of the next two sections parallels this section (excluding the present subsection, 4.4). We discuss Panel B (time horizon) of Figures 2 and 3 in Section 5, and Panel C (social circles) in Section 6.

5. Weights on Time Horizons

Few papers that use SWB data explicitly discuss which intertemporal preference information they capture. The exceptions make different assumptions including, at times, regarding the *same* SWB question; some assume that SWB data capture flow utility (e.g., Blanchflower and Oswald, 2004; Finkelstein, Luttmer, and Notowidigdo, 2013), while others assume that they capture forward-looking (or lifetime) utility (e.g., Aghion, Akcigit, Deaton, and Roulet, 2016).

To obtain evidence about the time period over which our respondents evaluated their situation when answering the SWB question, we asked them:

When you answered the [Life Satisfaction] Question, did you evaluate your situation as it is right this moment or over a longer period of time, in the past or in the future? To what extent did you evaluate your situation...

followed by ten sliders, in the same order, labeled from “Right this moment (while answering the survey)” to “Over your entire life, including your expectations for the future,” and followed by “Other (please specify).”

An SWB question that captured flow utility, $u(\mathbf{c}_\tau)$, would have the respondent evaluate her situation in the “present period” $\tau = t$.¹⁶ Depending on the economic application, the theoretical construct “present period” (or “period t ”) may be interpreted as including different possible time intervals around the moment of answering the SWB question, from a few minutes (e.g., in a laboratory experiment) to many years (e.g., in a lifecycle model). In principle, we could define a period’s length and include a slider for every period since the respondent’s birth. In practice, in order to keep the number of sliders reasonable and the response options intuitive, we instead opted for a limited number of naturally parsed periods, of different lengths. Thus, to shed light on whether the different SWB questions capture something that resembles a flow-utility concept and, if so, of what length, our survey question has sliders labeled “Right this

¹⁶ Here and in our theoretical framework in Section 2, to keep things simple, we write flow utility as a function of consumption in the current period, but our framework could be extended to allow flow utility to depend also on past consumption or expectations about future consumption, e.g., due to habit formation (for recent analyses, see Havranek, Rusnak, and Sokolova, 2017; Zhou, 2020), reference-dependence on past or future reference points (for a recent review, see O’Donoghue and Sprenger, 2018), or utility from memory or anticipation (e.g., Elster and Loewenstein, 1992; Morewedge, 2015). Because of the wording of our introspective question (“...did you evaluate your situation as it is right this moment or over a longer period of time...”), we believe that even if past consumption or expectations about future consumption affect flow utility, as long as SWB captures only flow utility, then respondents would report that they evaluate their situation in the present period. We similarly believe that respondents would report that they evaluate their situation in the present period if they evaluate their situation relative to their life in the past or to an important past event, as found by Ross, Eyman, and Kishchuk (1986), Ralph, Palmer, and Olney (2011), and Jungthaenel et al. (2018). To help ensure that respondents indeed report in this way, in our second survey we give them more explicit instructions, training, and feedback on this specific point; see Section 7.

moment,” “Today” and “In the last few [days]/[months]/[years]” (three different sliders, in this order).¹⁷

An SWB question that captured forward-looking or lifetime utility U_t would have the respondent evaluate their situation not only in the present but also in all future periods (as expected at t) and, for lifetime utility, also in all past periods. To capture various possibilities, our survey question includes sliders labeled “In the next few [months]/[years]” (in this order, two different sliders that may also capture an extended “present period” interpretation), as well as “Entire life so far” and “Entire life including your expectations for the future.” A pure measure of lifetime utility should put the most weight on this last timeframe.

5.1. General patterns

Figure 2 Panel B reports the mean weights respondents assigned to each time period, by SWB question. The standard error for each data point is roughly 1.7. The range of mean weights for time periods, between around 50 and around 70, is narrower than the range for life domains discussed in the previous section (roughly 35 to 75), and lies entirely to the right of the scale’s midpoint of 50. We interpret this to mean that for all the SWB questions, on average, respondents put positive weight on all the time periods—implying that none of the SWB questions cleanly captures flow utility or forward-looking utility.

At the same time, although the SWB questions fit our formal definition of lifetime utility (which merely requires positive weight on all periods of life), they do not correspond to a

¹⁷ We let these naturally parsed periods overlap. Alternatively, we could have eliminated overlap by replacing “Today” with “Today, excluding this moment”; replacing “In the last few days” with “In the last few days excluding today”; etc. We decided against adding these explicit exclusions because we worried that respondents would find them cumbersome and confusing. In our second survey, we did add these exclusions, together with detailed instructions to prevent confusion; see Section 7.

plausible version of lifetime utility. For six of the eight questions (the exceptions are Dealing Well and Meaning & Value, discussed below), “Right this moment” and “Today” rank higher than “Entire life so far” and “Entire life including your expectations for the future”; and for all questions, “Right this moment” ranks higher than “Today.” Even with present-biased time preferences, it is implausible that someone’s preferences would put more weight on a few minutes in the immediate present, or even on the rest of one’s day, than on one’s entire life.

5.2. Comparing across SWB questions

Despite the narrower range of mean responses for the time-horizon questions, we observe substantial differences across the eight SWB questions. Correlations (which ranged from 0.89 to 0.99 for the eight domain vectors) range from -0.17 to 0.96 , with median = 0.52 (Web Appendix C). The rightmost graph highlights three notable examples. At one extreme, the Personal Well-Being question gets the highest weight for “Right this moment” and “Today” and the lowest weight for “Entire life so far” and “Entire life including your expectations for the future,” making it the most flow-like among the eight questions—something we did not anticipate when formulating this question.

At the other extreme, the Dealing Well question, which explicitly asks about the past, gets the lowest weight for both “Right this moment” and “Today,” the highest for “Entire life so far,” and second-highest for both “Entire life including your expectations for the future” and “Last few years.” This profile is negatively correlated with those of six of the other SWB questions, and again suggests that respondents react in sensible ways to the wording of both the SWB question in the beginning of the survey and the introspective questions that follow it.

A third distinctive pattern is offered by the Meaning & Value question, with relatively high weights on “Right this moment” and “Today,” *and* on “Entire life including expectations” and “Entire life so far.” This unique combination (correlated 0.10–0.74 with other profiles) does not cleanly correspond to any utility notion we are aware of.

As seen most clearly in the center graph, the three standard SWB questions cannot, for the most part, be distinguished from each other in their time-horizon weights (correlations among the three are 0.94–0.96). All three get more weight on “Right this moment” and “Today” than on other time periods, with the Happiness question perhaps more so than the others. Overall, none of the three shows a pattern consistent with forward-looking or lifetime utility, nor do they exhibit the more flow-like pattern of Personal Well-Being.

In summary, we read our findings in this subsection as cautionary yet hopeful. On the one hand, the three traditional SWB questions appear not to have time profiles that cleanly capture flow, forward-looking, or lifetime utility. On the other hand, respondents react to the wording of SWB questions in sensible ways, suggesting that changing question wording may be effective at directing respondents towards a desired timeframe. Indeed, building on these findings, in Section 7 below we show that in our second survey, a happiness question that explicitly asks about “yesterday” appears to reasonably capture flow utility (with period length of one week or below).

5.3. Comparing across respondents

Figure 3 Panel B again aggregates responses across the eight SWB questions, and reports means by age, sex, income, and employment status of labor-force participants. Interestingly, we find that men and the employed introspect about the SWB questions in a somewhat less flow-like way than women and the unemployed, respectively: they report putting more weight on their

entire life so far (with or without explicitly including future expectations) relative to the present. We again see these findings as cautionary, this time about SWB comparisons across these groups without explicitly taking into account the possibility, suggested by our data, that different groups may perceive the same SWB questions as asking about different time horizons. (We do not conduct statistical tests here because they are underpowered, but we do so in Section 7 for analogous analyses on data from our second survey, where the no-difference null is strongly rejected for some cross-group comparisons.)

6. Weights on Social Circles

Just as with intertemporal preferences, few SWB applications discuss which other-regarding preference information is captured by the SWB measure. To obtain evidence about whose well-being our respondents considered when answering the SWB question, we first asked them:

When you answered the [Life Satisfaction] Question, to what extent did you evaluate your own, personal situation relative to evaluating the situation of a larger group that includes you and others?

A single slider, with a default initial value at the midpoint, was labeled “Personal situation” on its left end, and “Larger Group” on its right end. Respondents who allocated a positive weight to the latter (i.e., respondents who did not move the pointer all the way to the left), saw a follow-up screen with a more detailed set of sliders. They were asked:

When you answered the [Life Satisfaction] Question, to what extent did you evaluate the situation of ...

followed by eight sliders, in fixed order, labeled “Yourself,” “Your immediate family (parents, children, siblings, spouse),” “Other relatives,” “Your friends,” “Your community,” “Your country,” “The world,” and “Other (please specify).” As with our time-horizon question, while in principle we could have included a slider for every person in the world, in practice we opted for a limited number of naturally parsed groups.

6.1. General patterns

The “Larger group” row of Figure 2 Panel C shows that for all SWB questions, respondents allocated, on average, less weight to “Larger group” than to “Personal situation.” The mean weight varies from just below 30 (out of 100) for Happiness to above 40 for Family Well-Being. At the same time, as shown in the “%(Larger group > 0)” row, in all SWB questions a large majority of respondents allocated at least some (non-zero) weight to “Larger group,” ranging from 75% of respondents for Happiness to 90% for Family Well-Being. These findings suggest that for most respondents, none of the SWB questions is purely a measure of self-centered well-being. Formally, the $\tilde{\lambda}_k$'s for $k \neq 0$ are not all 0. We cannot draw strong conclusions from these data alone, however, since noise in responses would drive mean weights away from zero.

The rest of the rows show the results for the follow-up screen that was presented to the respondents who allocated positive weight to “Larger group.”¹⁸ Across SWB questions, the range

¹⁸ In the remainder of Section 6, we focus on analyzing responses to this follow-up screen. As explained above, it was not presented to respondents who gave 0 weight to “Larger Group” in the initial screen. In addition, due to a coding error, it was also not presented to respondents who did not move the slider on the initial screen from its default value at the midpoint between “Personal situation” and “Larger Group.” In Web Appendix B we combine, at the individual level, responses from the two screens (the initial Personal-situation-vs.-larger-group screen and the follow-up social-circles screen), and show that results remain very similar across alternative specifications, including specifications that include all respondents by imputing values for the missing follow-up-screen responses. (Our second survey uses an alternative design; see Section 7.)

of weights assigned to the response categories, 35 to 80, is wider than the range observed for life domains and time horizon. Correlations are higher too, ranging from 0.93 to 1.00. For each of the eight SWB questions, “Yourself” was allocated the most weight—always above 70—with “Immediate Family” an unambiguous second—always above 60, with the single exception of Personal Well-Being discussed below. All other social categories were allocated less weight, with relatively little variation across them. These findings more strongly rule out the hypothesis, mentioned above, that some SWB questions elicit a fully self-centered well-being notion: all eight SWB questions seem to contain a substantial immediate-family component.

6.2. Comparing across SWB questions

As seen in the rightmost graph, the comparison between Personal Well-Being and Family Well-Being again suggests that respondents react sensibly to SWB-question wording. Nearly identical except for the mention of family, the two questions indeed virtually coincide on all sliders other than “Yourself” and “Immediate Family”—which get weights of 80 and 60 for Personal Well-Being, compared with 73 and 72 (not statistically distinguishable) for Family Well-Being. Formally, while we do not know the utility weights λ_k , these findings suggest that for Family Well-Being, either $\tilde{\lambda}_k = 1$ or $\sum \tilde{\lambda}_k = 1$ for k corresponding to family members—depending on whether a respondent interprets “Immediate Family” as referring to each member or to their sum.

In the center graph, the three traditional SWB questions appear similar to each other, with nearly identical profiles for Life Satisfaction and Happiness. For Ladder, respondents assign a slightly lower weight on everything other than “Yourself,” a pattern that we did not anticipate. As with time horizon, these three questions appear to occupy a middle ground among the eight

questions, neither centered more on self nor on the larger social circles. Overall, they appear consistent with other-regarding preferences: $\tilde{\lambda}_k > 0$ for k corresponding to family members and, possibly, also wider social circles.

Finally, for Meaning & Value, and to a lesser extent for Options & Possibilities, respondents assign higher weights to individuals outside the immediate family (leftmost graph). This pattern is consistent with these SWB questions capturing more of the other-regarding components of preferences. It is also consistent with the finding from Section 4.2 above that these SWB questions are associated with higher weights on the domains “Volunteering, activism” and “Family life & relationships.”

6.3. Comparing across respondents’ sociodemographics

Averaging across SWB questions, Figure 3 Panel C shows few differences across sociodemographic groups in the ordinal ranking of mean weights. Looking at mean-weight *magnitudes*, women, older, and unemployed respondents are higher or equal on one’s self and immediate family and lower on wider social circles relative to men, younger, and employed respondents. However, for the sex and age comparisons—but not for employment comparison—these differences could also reflect cross-group differences in scale-use. Indeed, as Panel C shows, men’s and the young’s mean social-circles weights are consistently closer to the scale midpoint, 50.

In the next section we use our second survey’s alternative design—which, among other things, aims to reduce scale-use heterogeneity—and significantly larger sample size to formally and systematically test (and reject) the null of no cross-group differences in mean weights.

7. Conceptual Replication: Second Survey

To test the robustness and replicability of our conclusions from the first survey, we conducted a second survey on September 24–October 7, 2022. To assess robustness, we varied many of its design details relative to our first survey. In addition, it included several new design elements intended to address potential limitations of the first survey. In this section we summarize its design and main findings. For survey screenshots, see Web Appendix F.

7.1. Survey design and sample

Our second survey differs from our first survey in the following ways: it focuses on only five SWB questions; the sample is four times larger (per SWB question) and more representative of the U.S. population; the survey asks each respondent all five SWB questions and only two introspective questions about each; the introspective questions require allocation of a fixed budget across (a smaller set of) mutually exclusive response categories; and the survey includes extensive instructions and comprehension checks.

The five SWB questions include four from Section 3.1: two that are most commonly used in applied work—Happiness and Life Satisfaction—and two new ones that, as discussed above, appear promising and at the same time provide a clean test of social-circle wording variation—Personal Well-Being and Family Well-Being. To these, we added a fifth question:

Happiness Yesterday:

Overall, how happy did you feel yesterday? Please give a number between 0 (extremely unhappy) and 10 (extremely happy): _____

We included this question both because this wording is used in the U.K.'s Office for National Statistics Integrated Household Survey (based on the recommendation by Dolan and Metcalfe, 2012) and because its comparison with the original Happiness question provides an additional test—in this case regarding how wording affects time-horizon interpretation.

Our sample is $N = 1,497$ online respondents, recruited on Prolific to try to match the adult U.S. population on age, sex, and ethnicity (see Web Appendix A). While our respondents roughly match the population on sex, marital status, household size, and fraction that is White, they are older, more educated, higher income, and more Northeast and less West.

Unlike in our first survey, each respondent in our second survey is asked *each* of the five SWB questions, randomly ordered; like Figure 1, Figure 4 reports histograms summarizing responses to the SWB questions, by SWB question and pooled. Each SWB question is followed by two introspection questions (also randomly ordered), which use a newly designed graphical tool. Respondents use the mouse and/or keyboard to distribute 100 points across mutually exclusive time-horizon categories on a timeline or social-circles categories on a diagram. The tool is introduced step-by-step with extensive instructions, and the timeline and diagram are dynamically constructed using animation. Respondents go through several examples, hands-on training in using the tool, and detailed instructions on how to interpret the introspection questions. The introspection questions themselves are substantially more detailed than in our first survey, and read as follows:

When answering the [**Happiness**]/[**Well-Being**]/[**Life Satisfaction**] **question**, [which social circles]/[what time periods] did you interpret the question as asking about?

Please use the tool below to answer this question.

Remember:

- You should report how you interpreted the question when answering it and **not** how you think, after the fact, you should have interpreted it.
- We are only interested in what you interpreted the question as asking about and **not** what comparisons you made.
- We are only interested in what you interpreted the question as asking about and **not** what else may have affected your answer.

The tool consists of a bar at the right-hand side of the screen that initially contains the 100 points to be distributed into the different category bars. In the time-horizon question, seven category bars are assigned to labeled intervals on the timeline, with “The Past Week” in its center, “Rest of Last Month” immediately to the left, followed by “Rest of Last Year” and “Further in the Past” and, to the right, “The Month After,” “Rest of Next Year,” and “Further in the Future.” In the social-circles question, five bars are assigned to concentric circles labeled “You,” “Your Family,” “Friends & Coworkers,” “Your Community,” and “The World.”

The three bullets under “Remember” in the introspection question above are designed as short reminders, summarizing three important question-interpretation points. Each of the three is explained in detail earlier in the instructions, along with a spelled-out example, a comprehension question, and detailed feedback. The first emphasizes that if a respondent interpreted the SWB question, *at the time of answering it*, as referring, for example, to only herself, then she should not put weight on her family even if, upon further reflection, she now thinks she should have interpreted things differently. The second and third emphasize that if a respondent interpreted the SWB question as referring, for example, to only the past week—arguably, a flow-utility interpretation—then she should not put weight on, say, the past year even if she evaluated her well-being in the past week *relative* to the past year, or relative to past expectations, and even if

her well-being in the past week was *affected* by earlier events (or their memory). Similarly, she should not put any weight on the future, even if her well-being in the past or present was affected by anticipation, expectations, or beliefs regarding the future, unless she interpreted the SWB question as actually asking about her well-being also the future—as she would under lifetime- or forward-looking-utility interpretations.

Since respondents answer the five SWB questions in random order, this survey combines a between-subjects design (by analyzing only the first SWB question each respondent saw) with a within-subject design (by analyzing all data); we discuss below within-subject results (between-subjects results are qualitatively similar, but standard errors are larger). Before the second and third SWB questions, respondents answer attention-check questions (90% of respondents answered both attention questions correctly). To avoid potential demand effects, unlike the first survey, the current SWB question does not appear on the introspection-question screens, and is referred to without using words that could potentially imply a social circle or time horizon; hence both Family and Personal Well-Being questions are referred to in the introspection question above as “the Well-Being question,” and both Happiness and Happiness Yesterday questions are referred to as “the Happiness question.” The survey ends with demographic questions. It was programmed using oTree (Chen et al., 2016). Median completion time was 19 minutes.

7.2. Survey results: comparing across SWB questions

Figure 5 Panel A shows that while survey design and/or sample details matter, the main conclusions are the same across the two surveys. Looking at the weights on time horizons, Happiness Yesterday, which is explicit about time (“yesterday”) and was not included in the first

survey, is the only SWB question in the second survey that appears consistent with a rather clean flow-utility interpretation: it has 88% of the weight, on average, on the past week. In contrast, all four other questions from the first survey, including the commonly used Happiness and Life Satisfaction questions, have 47–58% on the past week, 15–17% on the rest of last month, 10–14% on the rest of last year, 6–9% on further in the past, and 11–14% on the future—suggesting a mix of mostly flow and backward-looking utility (or, alternatively, flow utility with a mix of lengths of the current period t). As in our first survey, Happiness and Personal Well-Being appear more flow-like than Life Satisfaction and Family Well-Being.

Looking at social circles, with the expected exception of Family Well-Being—the only SWB question explicit about family—the SWB questions again look rather similar to each other. While, unlike in the first survey, the wider circles of country and world get essentially no weight under the second-survey’s design constraint of 100 total points, the four SWB questions get overwhelmingly most—but far from all—weight on self, then family, then friends and beyond. Again, the commonly used Happiness and Life Satisfaction do not appear cleanly self-centered—with only 77% and 72% of the weight on self, respectively—and have a significant family component. Again Personal Well-Being looks the most self-centered, this time together with Happiness Yesterday: 84% and 82% on self, respectively. Finally, and also replicating the earlier results, Family Well-Being is dramatically less self-centered, with 54% on the weight on family and 40% on self, compared with 10% and 84%, respectively, for Personal Well-Being.

Statistically, as the tight (indeed, hardly noticeable) standard-error bars in Figure 5 suggest, the five SWB questions are almost always distinguishable from each other on both time profiles and social circles. For example, testing for equality of average percentage of points respondents put on past week or on self, seventeen of the twenty pairwise equality-of-means t -

tests are rejected with p -value < 0.001 , with the exceptions being Happiness versus Personal Well-Being (2 percentage-point difference on past week, $p = 0.009$), Happiness Yesterday versus Personal Well-Being (1.5 point difference on self, $p = 0.016$), and Life Satisfaction versus Family Well-Being (1 point difference on past week, $p = 0.35$).

7.3. Survey results: comparing across respondents' sociodemographics

Finally, Figure 5 Panel B qualitatively replicates the general finding that different demographic groups may interpret SWB questions differently, although the specific findings differ from those of the first survey, perhaps due to the differences in survey design and sample. Looking at the same set of eight comparisons (across four sociodemographic characteristics) we originally analyzed in the first survey, and reporting everything that is statistically suggestive, in the second survey data we find that the unemployed are more self-centered than the employed (by 4.5 percentage points on self, $p = 0.006$), the young are more flow-like than the old (by 10 points on the past week, $p = 2.50 \times 10^{-10}$, ages 18–35 versus 56+; ages 36–55 are exactly in between), and low-income respondents are possibly both more self-centered and more flow-like than high-income ones (by 4 and 6 points, $p = 0.019$ and 1.6×10^{-6} , income below \$40k versus \$80k and up; income \$40-80k are in between).

Intriguingly, however, when we focus on the time-explicit Happiness Yesterday (not shown in the figure), we find much weaker evidence of the above time-horizon differences. In particular, the largest difference above—weight on the past week between young and old—is a substantial 12 points ($p = 1.72 \times 10^{-7}$) in the time-horizon-ambiguous Happiness question, but only 3 points ($p = 0.072$) in its time-horizon-explicit counterpart, Happiness Yesterday. (For the cross-SWB-question difference between these two differences, $p = 1.28 \times 10^{-4}$.) This finding

suggests that making SWB questions less ambiguous could, in addition to moving them closer to desired utility notions *on average*, also reduce their interpretation *variance*.

8. Discussion and Concluding Remarks

The now-standard SWB questions that are regularly asked on large-scale social surveys were originally designed during the 1920s through 1970s by marriage researchers, education and personality psychologists, mental-health epidemiologists, gerontologists, and social-indicator researchers (Angner, 2011). These researchers had a variety of notions they intended to measure with these questions—but none designed their questions with the utility notions that economists have in mind when they use SWB data today.

In this paper, we use two introspective surveys to evaluate the extent to which responses to existing SWB survey questions might correspond with any of the utility notions researchers assume they represent. We find that, first, according to respondents' reported introspective weights, almost none of the SWB measures we studied, including those based on standard happiness, life satisfaction, and ladder questions, have the time profile of flow utility, forward-looking utility, or lifetime utility; happiness yesterday is a promising exception, further discussed below. Second, none of the measures corresponds to purely self-centered utility; instead, each incorporates concern for others, particularly one's family. At the same time, respondents' weights consistently react as expected to differently worded SWB questions, a point we return to shortly.

We also aimed to test the extent to which there is heterogeneity across respondents in the time horizons and social circles captured by their responses. Across sociodemographic groups,

we find both that such heterogeneity exists and, importantly, that this heterogeneity itself may also react as expected to differences in wording.

While our paper is primarily aimed at addressing assumptions that SWB data capture some component of preferences—as assumed in many applications—our finding of heterogeneity in the weights used to construct SWB responses is problematic even if the goal of policy is taken to be happiness (as in, e.g., Layard, 2005). If some individuals report a mix of their own happiness with their family’s whereas others report their own, or if some individuals report their momentary happiness whereas others report a longer-term average, aggregating these disparate data into a normative measure of welfare is challenging.

Our results point to some readily applicable practical advice for researchers. Our advice depends on the researchers’ latitude to shape the survey data they analyze. To users of existing SWB data, we caution against automatically interpreting SWB analyses as measures of the standard utility notions. Researchers should keep this caution in mind when drawing scientific and policy conclusions from such analyses. To researchers who add their own SWB question to an ongoing survey, we recommend tweaking the standard wording of a question if doing so can bring it more in line with the utility notion the responses will be used to represent; most straightforwardly, researchers aiming for a flow utility measure should ask about SWB during some explicit recent time interval, e.g., yesterday or in the past week, as many surveys already do (e.g., the Office for National Statistics Integrated Household Survey and the Gallup-Healthways Well-Being Index). Indeed, among the nine SWB questions investigated in this paper, our respondents report that one such question—“Overall, how happy did you feel yesterday?”—comes the closest to eliciting self-centered flow utility; and our newly phrased Personal Well-Being question—“On a scale from 0 to 10, how would you rate your overall

personal well-being?”—while still far from cleanly eliciting flow utility, comes as close to eliciting self-centered utility. These short and simple questions may be promising points of departure for further tweaking.

To researchers who can add multiple questions to a survey or are designing their own survey (or can do an auxiliary survey on a similar sample), we additionally suggest that it may be useful to include introspective questions like ours. These can be used to shed light on how successfully the SWB question gets respondents to think about their response in a way consistent with the desired utility notion or to diagnose heterogeneity in weights across respondents.

When an analysis involves regressing SWB on sociodemographics, introspective questions like ours may also provide a means of assessing how robust the conclusions are to heterogeneity in SWB-question interpretation across respondents. Specifically, researchers could assess how much the estimated regression coefficients change with and without controlling for the weights respondents report to the introspective questions. In Web Appendix B, we formalize the theoretical framework underlying such an analysis and conduct it using our results from both surveys. We indeed find that when we control for differences across respondents in their social-circle weights (but not their time-horizon weights), the coefficients from a regression of SWB on sociodemographics change, in some cases substantially. Moreover, since at the respondent level, our measures of the time-horizon and social-circle weights are likely noisy, our analyses likely understate the degree to which the magnitudes of sociodemographic comparisons of SWB are affected by the differences in weights. We therefore advise against relying heavily on the magnitudes of coefficients from SWB regressions for policy purposes (as advocated by, e.g., Bronsteen, Buccafusco, and Masur, 2013, and Frijters, Clark, Krekel and Layard, 2020).¹⁹ We

¹⁹ Prior work has led to this same recommendation based on comparing the MRSs implied by SWB measures to the MRSs implied by choice (Benjamin, Heffetz, Kimball, and Rees-Jones, 2014).

caution, however, that even for the signs of coefficient estimates in sociodemographic comparisons of SWB, which are much more robust to our controls for heterogeneity than the magnitudes, interpreting these comparisons relies on additional assumptions that we have not tested in the present paper, such as sufficiently similar uses of the SWB response scales across the groups (see, e.g., Benjamin, Cooper, Heffetz, and Kimball, 2020).

More broadly, we believe that our methodology of asking introspective questions could be useful in studying other aspects of how survey respondents answer SWB questions. Indeed, while this paper focuses on results pertaining to how respondents weight different life domains, time horizons, and social circles, our first survey included additional introspective questions about how respondents chose a particular 0–10 number when answering the SWB question (e.g., whether they were comparing to some absolute scale or a scale relative to other people, their life in the past, or their goals). We have not analyzed these data in detail, but will share them upon acceptance at a journal—as we will our code from both surveys, including that of the graphical interface from our second survey. We hope they will be useful for future research.

References

- Adler, Matthew D.** 2012. *Well-Being and Fair Distribution: Beyond Cost-Benefit Analysis*. New York: Oxford University Press.
- Adler, Matthew D.** 2013. “Happiness Surveys and Public Policy: What’s the Use?” *Duke Law Journal*, 62(8): 1509–1601.
- Adler, Matthew D.** 2019. *Measuring Social Welfare: An Introduction*. New York: Oxford University Press.
- 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.
- Aghion, Philippe, Ufuk Akcigit, Angus Deaton, and Alexandra Roulet.** 2016. “Creative Destruction and Subjective Well-Being.” *American Economic Review*, 106(12): 3869–3897.

- 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.
- Allcott, Hunt, Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow.** 2020. "The welfare effects of social media." *American Economic Review*, 110(3), 629–676.
- Andreoni, James.** 1989. "Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence." *Journal of Political Economy*, 97(6): 1447–1458.
- Angner, Erik.** 2011. "The evolution of eupathics: The historical roots of subjective measures of wellbeing." *International Journal of Wellbeing*, 1(1): 4–41.
- 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.
- Bénabou, Roland and Jean Tirole.** 2006. "Incentives and Prosocial Behavior." *American Economic Review*, 96(5): 1652–1678.
- Benjamin, Daniel J., Kristen Cooper, Ori Heffetz, and Miles S. Kimball.** 2019. "A Well-Being Snapshot in a Changing World." *American Economic Review (Papers and Proceedings)*, 109: 344–349.
- 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.
- 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.
- 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.
- 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.
- 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.
- Blanchflower, David G.** 2020. "Is happiness U-shaped everywhere? Age and subjective wellbeing in 145 countries." *Journal of Population Economics*, 34: 575–624.
- 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.
- Blanchflower, David G., and Andrew J. Oswald.** 2008. "Is well-being U-shaped over the life cycle?" *Social Science & Medicine*, 66: 1733–1749.
- Bronsteen, John, Christopher Buccafusco, and Jonathan S. Masur.** 2013. "Well-Being Analysis vs. Cost-Benefit Analysis." *Duke Law Journal*, 62: 1603–1689.

- Cantril, Hadley.** 1966. *The pattern of human concerns*. New Brunswick, N.J.: Rutgers University Press.
- Chen, Daniel L, Martin Schonger, and Chris Wickens.** 2016. “oTree—an open-source platform for laboratory, online, and field experiments.” *Journal of Behavioral and Experimental Finance*, 9: 88–97.
- Cummins, Robert A.** 1996. The domains of life satisfaction: An attempt to order chaos. *Social Indicators Research*, 38(3): 303–328.
- Deaton, Angus, Jane Fortson, and Robert Tortora.** 2010. “Life (Evaluation), HIV/AIDS, and Death in Africa.” In *International Differences in Well-Being*, ed. Ed Diener, Daniel Kahneman and John Helliwell, 105–137. UK: Oxford University Press.
- 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.
- Decancq, Koen, Marc Fleurbaey, and Erik Schokkaert.** 2015. “Happiness, equivalent incomes, and respect for individual preferences.” *Economica*, 82, 1082–1106.
- Dunn, Elizabeth W., Lara B. Aknin, and Michael I. Norton.** 2008. “Spending Money on Others Promotes Happiness.” *Science*, 319(5870): 1687–1688.
- Easterlin, Richard A.** 1995. “Will raising the incomes of all increase the happiness of all?” *Journal of Economic Behavior & Organization*, 27(1): 35–47.
- Easterlin, Richard A.** 2006. “Life cycle happiness and its sources: Intersections of psychology, economics, and demography.” *Journal of Economic Psychology*, 27(4): 463–482.
- Edgeworth, Francis Y.** 1881. *Mathematical Psychics*. New York: Augustus M. Kelley.
- Elster, Jon, and George Loewenstein.** 1992. “Utility from memory and anticipation.” In *Choice over time*, ed. G. Loewenstein and J. Elster, 213–234. New York: Russell Sage.
- 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.
- 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.
- Fleurbaey, Marc, and Hannes Schwandt.** 2016. “Do People Seek to Maximize Their Subjective Well-Being – And Fail?” Manuscript.
- Frank, Robert H.** 1985. *Choosing the Right Pond: Human Behavior and the Quest for Status*. Oxford University Press.
- 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.
- Glaeser, Edward L., Joshua D. Gottlieb, and Oren Ziv.** 2016. “Unhappy Cities.” *Journal of Labor Economics*, 34(2): 129–182.
- Glaeser, Edward L. and Charles Redlick.** 2009. “Social Capital and Urban Growth.” *International Regional Science Review*, 32(3): 264–299.

- Gruber, Jonathan and Sendhil Mullainathan.** 2002. “Do Cigarette Taxes Make Smokers Happier?” NBER Working Paper No. 8872.
- Gruber, Jonathan, and Sendhil Mullainathan.** 2005. “Do Cigarette Taxes Make Smokers Happier?” *B.E. Journal of Economic Analysis and Policy*, 5(1).
- Hausman, Daniel M.** 2012. *Preference, Value, Choice, and Welfare*. New York, N.Y.: Cambridge University Press.
- Havranek, Tomas, Marek Rusnak, and Anna Sokolova.** 2017. “Habit formation in consumption: A meta-analysis.” *European Economic Review*, 95: 142–167.
- Jackson, Matthew O., and Leeat Yariv.** 2015. “Collective dynamic choice: the necessity of time inconsistency.” *American Economic Journal: Microeconomics*, 7(4): 150–178.
- 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.
- 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.
- Kahneman, Daniel, Peter P. Wakker, and Rakesh Sarin.** 1997. “Back to Bentham? Explorations of experienced utility.” *Quarterly Journal of Economics*, 112: 357–406.
- Kapteyn, Arie, James P. Smith, and Arthur van Soest.** 2009. “Life satisfaction.” IZA Discussion Paper No. 4015.
- Kimball, Miles, and Robert Willis.** 2006. “Happiness and Utility.” <http://www-personal.umich.edu/~mkimball/pdf/uhap.pdf>
- Laibson, David.** 1997. “Golden Eggs and Hyperbolic Discounting.” *Quarterly Journal of Economics*, 112(2): 443–477.
- Layard, Richard.** 2005. *Happiness: Lessons From A New Science*. London: Penguin Books.
- 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.
- Millner, Antony, and Geoffrey Heal.** 2018. “Time consistency and time invariance in collective intertemporal choice.” *Journal of Economic Theory*, 176: 158–169.
- Morewedge, Carey K.** 2015. “Utility: Anticipated, Experienced, and Remembered.” In *Wiley-Blackwell handbook of judgment and decision making, Volume 1*, ed. Gideon Keren and George Wu, 295–330. New Jersey: Wiley-Blackwell.
- O’Donoghue, Ted, and Charlie Sprenger.** 2018. “Reference-Dependent Preferences.” In *Handbook of Behavioral Economics – Foundations and Applications, Volume 1*, ed. B. Douglas Bernheim, Stefano DellaVigna, and David Laibson. Oxford: Elsevier.
- Oswald, Andrew J.** 2008. “On the curvature of the reporting function from objective reality to subjective feelings.” *Economics Letters*, 100: 369–372.

- 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.
- Railton, Peter.** 1986. "Facts and Values." *Philosophical Topics*, 14(2): 5–31.
- 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.
- 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.
- Ross, Michael, Adele Eyman, and Natalie Kishchuk.** 1986. "Determinants of Subjective Well-Being." In *Relative Deprivation and Social Comparison: The Ontario Symposium, Volume 4*, ed. James M. Olson, C. P. Herman, and Mark P. Zanna. New Jersey: L. Erlbaum Associates.
- 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.
- Schwandt, Hannes.** 2016. "Unmet Aspirations as an Explanation for the Age U-shape in Wellbeing." *Journal of Economic Behavior and Organization*, 122: 75–87.
- Schwarz, Norbert and Fritz Strack.** 1999. "Reports of Subjective Well-Being: Judgmental Processes and Their Methodological Implications." In *Well-Being: The Foundations of Hedonic Psychology*, ed. Daniel Kahneman, Ed Diener, and Norbert Schwarz. New York: Russell Sage Foundation.
- 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.
- Stevenson, Betsey, and Justin Wolfers.** 2009. "The Paradox of Declining Female Happiness." *American Economic Journal: Economic Policy*, 1(2), 190–225.
- 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.
- 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.
- Szabó, Andrea, and Gergely Ujhelyi.** 2017. "Choice and Happiness in South Africa." *Economics Letters*, 155: 28–30.
- Van Praag, Bernard M., Paul Frijters, and Ada Ferrer-i-Carbonell.** 2003. "The anatomy of subjective well-being." *Journal of Economic Behavior & Organization*, 51(1): 29–49.
- WHR.** 2021. *The World Happiness Report*. <https://worldhappiness.report>
- Zhou, Jiannan.** 2020. "Survey Evidence on Habit Formation: Existence, Specification, and Implication." https://jiannanzhou.com/research/habit/Habit_Zhou_Paper.pdf

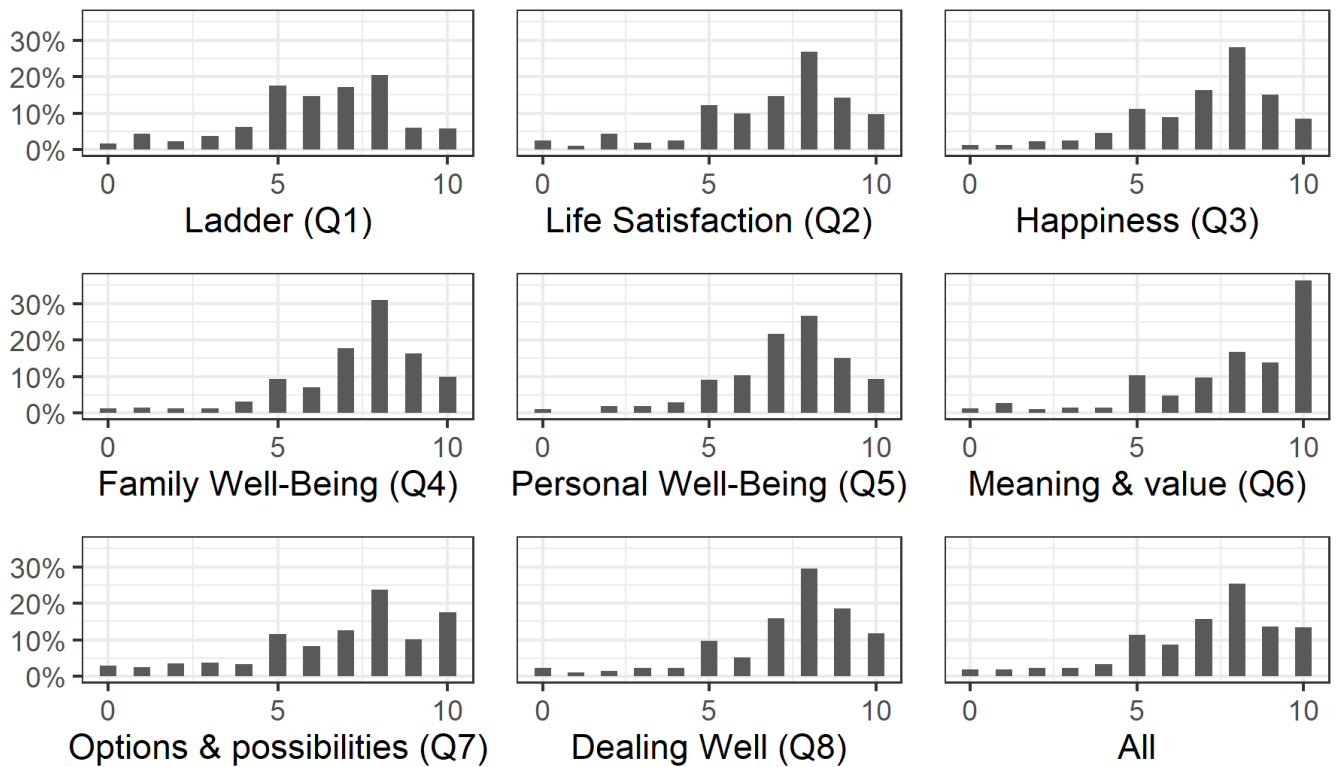


Figure 1. Response Histograms, by SWB Question (First Survey). Based on 3,040 observations (All) and on 359–397 observations (each of Q1–Q8). Two respondents entered non-integers; these were rounded to the nearest integers.

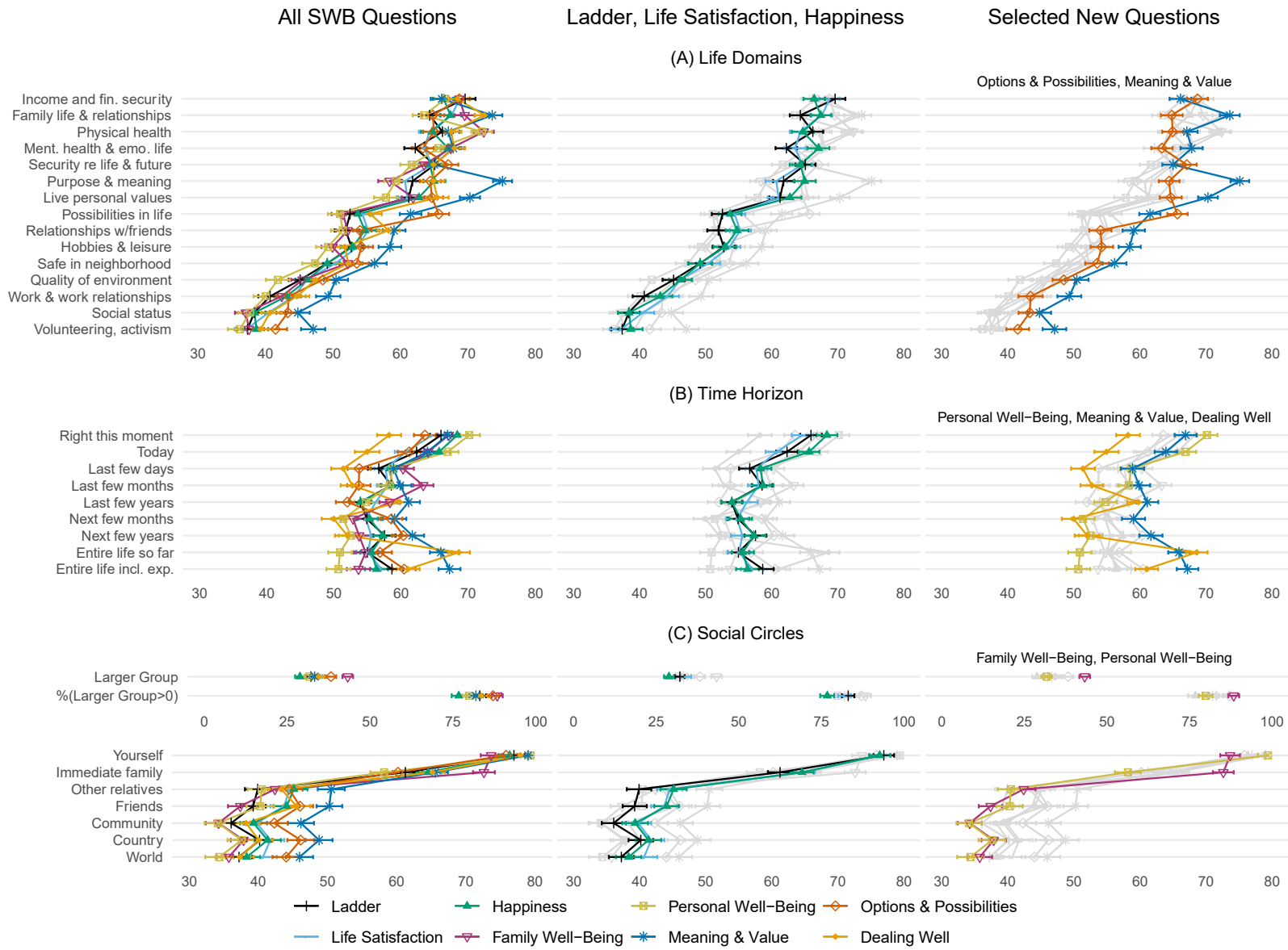


Figure 2. Reported Weight Placed, by SWB Question (First Survey). Based on 3,040 observations. Each row reports mean rating (0–100) by SWB question, other than “%(Larger Group > 0)” row, which reports percent of respondents who rated Larger Group above 0 (see text for details). “All SWB questions” column reports means/percent for all eight questions; “Ladder, Life Satisfaction, Happiness” column grays out all but these three widely used SWB questions; “Selected SWB questions” column grays out all but the two or three questions in the relevant mini-graph subtitle. Capped bars report standard errors.

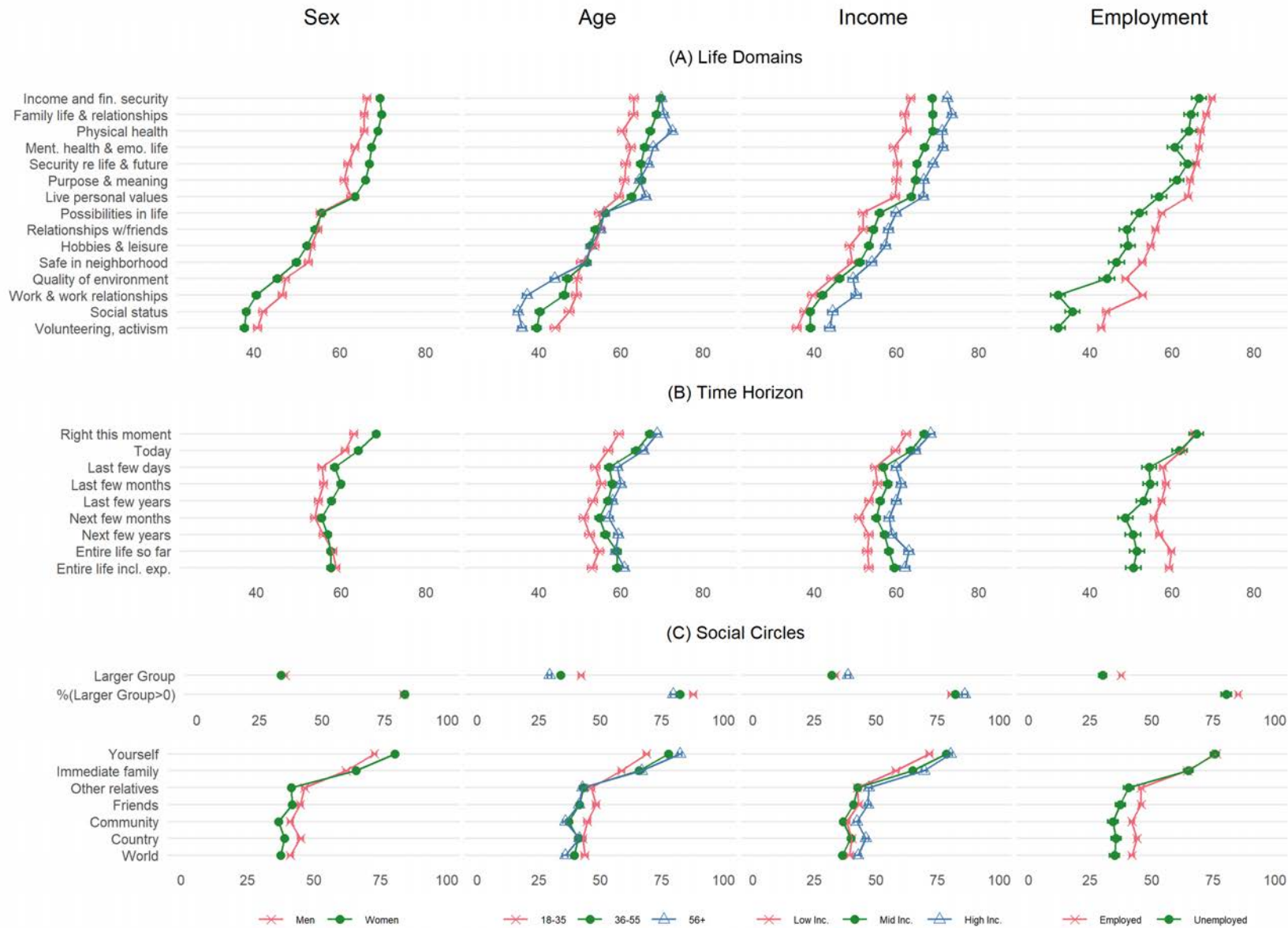
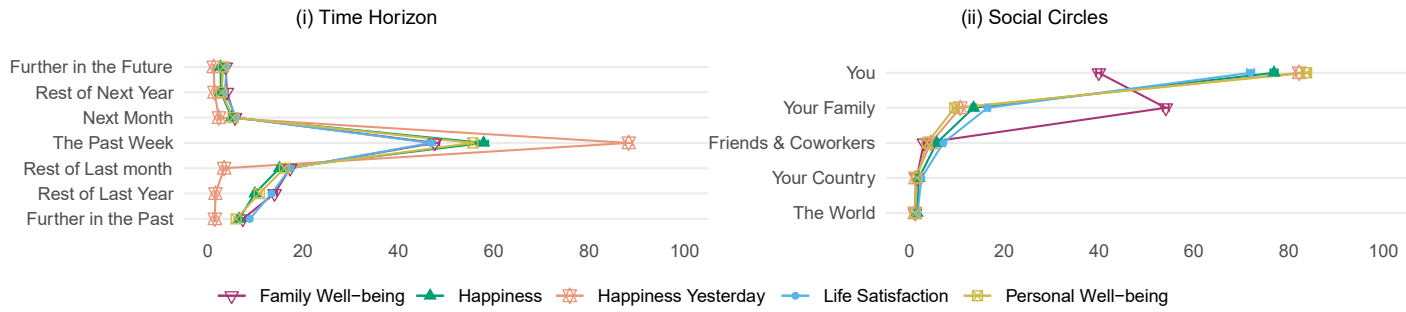


Figure 3. Reported Weight Placed, by Demographics (First Survey). Based on 3,040 observations (fewer when demographic information is missing; see Web Appendix A). Each row reports mean rating (0–100) by demographic group, other than “%(Larger Group > 0)” row, which reports percent of respondents who rated Larger Group above 0 (see text for details). The employment mini-graphs are based on only 1,590 observations (respondents not in the labor force are dropped). Capped bars report standard errors.



Figure 4. Response Histograms, by SWB Question (Second Survey). Based on 1,497 respondents, hence 1,497 observations for each of Q2–Q5 and Q9, and 7,485 observations for All.

(A) By SWB Question



(B) By Socio-Demographics

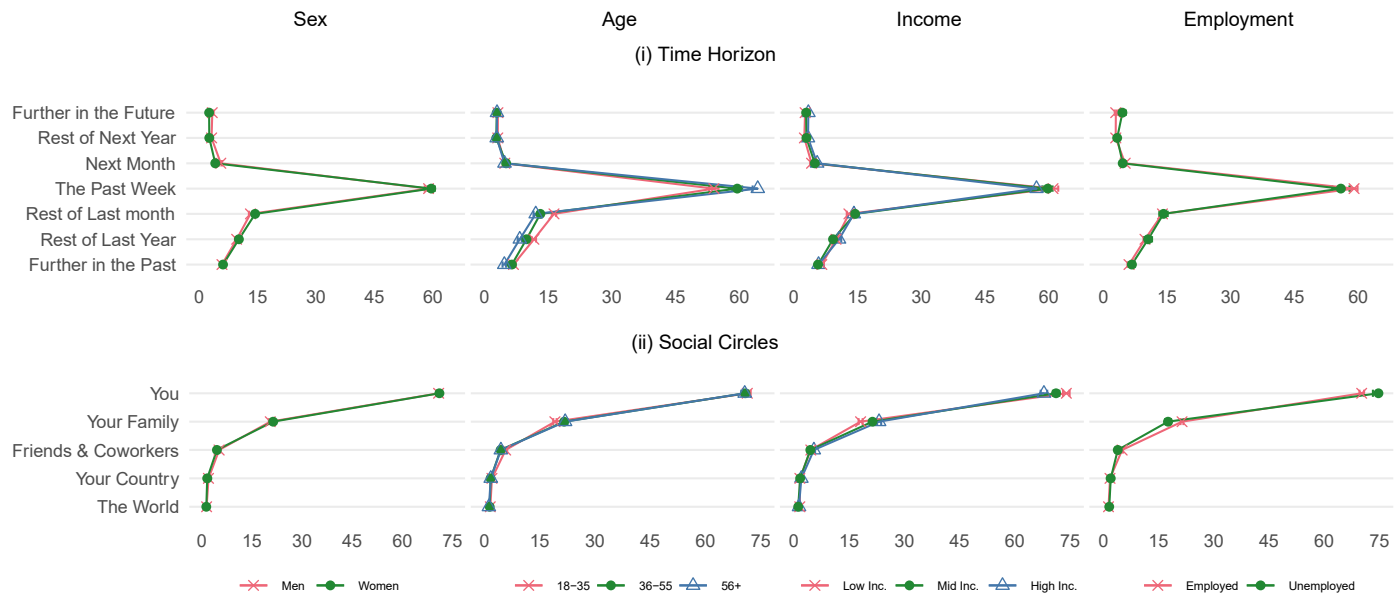


Figure 5. Reported Weight Placed (Second Survey). Based on 1,497 respondents, i.e., 7,485 observations (fewer when demographic information is missing; see Web Appendix A). Each row reports mean weight (0–100) by SWB question or by demographic group. The employment mini-graphs are based on only 1,181 respondents (those not in the labor force are dropped). Capped bars (often so tight that they are hardly noticeable) report standard errors, clustered at the respondent level.