What Do Happiness Data Mean? Theory and Survey Evidence*

Daniel J. Benjamin

Jakina Debnam Guzman

Marc Fleurbaey

Ori Heffetz

Miles Kimball

This Draft: May 16, 2021

First Incomplete and Preliminary Draft: May 31, 2014

Abstract

What utility notion—e.g., flow/lifetime, self/family-centered—do self-reported well-being (SWB) questions measure? First, we clarify the theoretical assumptions underlying existing applications regarding the (i) life domains, (ii) time horizons, and (iii) other-regarding preferences captured by SWB data. Second, we document inconsistency in assumptions across papers, sometimes using the same SWB dataset. Third, 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. We outline actionable suggestions for SWB researchers. [100 words]

JEL Classification: D69, D90, I31

Keywords: happiness, life satisfaction, subjective well-being, survey questions

* Benjamin: UCLA Anderson School of Management and UCLA David Geffen School of Medicine; NBER (e-mail: daniel.benjamin@gmail.com); Debnam Guzman: Department of Economics, Amherst College (e-mail: jguzman@amherst.edu); Fleurbaey: Paris School of Economics (e-mail: marc.fleurbaey@psemail.eu); Heffetz: S.C. Johnson Graduate School of Management, Cornell University; Bogen Family Department of Economics and Federmann Center for the Study of Rationality, The Hebrew University of Jerusalem; NBER (e-mail: oh33@cornell.edu); Kimball: Department of Economics, University of Colorado Boulder; NBER (e-mail: Miles.Kimball@colorado.edu). 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, and seminar participants at Colorado Boulder, London School of Economics, Paris School of Economics, Princeton, Stockholm Institute for Future Studies, Warwick, UC Berkeley, USC, and Louvain la Neuve. We are grateful to Ophir Averbuch, Joel Becker, Samantha Cunningham, Ofer Glicksohn, Arshia Hashemi, Aharon Haver, Yuezhou (Celena) Huo, Mattar Klein, Lev Maresca, 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 Woodrow Wilson School 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.

Web appendixes are available at the authors' websites.

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.

This paper makes two contributions. The first is a theoretically guided critical assessment of the literature. We call attention to widespread inconsistency across papers regarding the assumptions made about SWB data—in some cases, with different assumptions made about the very same data. Moreover, we show how the appropriate scientific and policy conclusions that can be drawn in typical applications often hinge on which, if any, utility-notion assumption holds. The second contribution is empirical. To provide some initial evidence relevant to a range of assumptions researchers make about SWB data, we conduct a survey where respondents introspect and report on how they construct their own answers to one of eight different SWB questions. Our conclusions are largely negative: respondents' reports do not neatly fit what would be predicted by any of the utility notions that researchers assume. However, we also find that variations in question wording have predictable effects on respondents' reports, which we view as a silver lining: it suggests that SWB questions could be designed that come closer to capturing a desired utility notion.

[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.)

Sections I and II constitute the first part of the paper. In Section I, we lay out a simple theoretical framework whose purpose is to clarify the assumptions typically made in the literature. 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.

In Section II, we consider the SWB literature in light of our theoretical framework. We begin by briefly reviewing work that investigates whether SWB captures the same utility notion as choices (actual or hypothetical). As mentioned above, that work finds deviations, 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. We then document how applications that use SWB data, while routinely assuming away (i), make varying assumptions about (ii) and (iii).

We show how in many common applications, researchers assume, explicitly or implicitly, that SWB data capture narrower utility notions than choice (e.g., flow, self-centered), and we illustrate how the scientific and policy conclusions that can be drawn often differ dramatically depending on which assumption is made. For example, in economic applications, researchers often run a regression of current SWB on current income and current consumption of a "good" such as employment (e.g., Clark and Oswald, 2002), and calculate the money value of the good as the ratio of coefficients. We show that the theoretical interpretation of this empirical ratio depends on which utility notion is captured by SWB. For example, if SWB measures flow utility, then the estimated "dollar cost of unemployment" is a one-time cost, whereas if SWB measures lifetime utility, then (under some simplifying assumptions) it is an *annual* cost—so the total cost

will be many times higher. We also illustrate how assumptions about the SWB measure matter for drawing inferences from SWB data about whether people are making mistakes (as in, e.g., Dunn, Aknin, and Norton, 2008) and for interpreting SWB's increase with age later in life (e.g., Blanchflower and Oswald, 2008; Stone, Schwartz, Broderick, and Deaton, 2010). We conclude from the first part of the paper that SWB researchers should be explicit and consistent about what utility notion they assume their SWB data capture, and should collect evidence to support that assumption.

Sections III–VI constitute the second part of our paper: in order to provide some initial empirical evidence on (i), (ii), and (iii), we analyze a survey in which we elicit respondents' introspections on how they constructed their SWB responses. In Section III, we describe the design of our 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 new 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 IV, 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 (IV.D), we discuss how we address limitations of the introspective methodology, with reference to examples from earlier in the section. In Sections V and VI 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 IV-VI, 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 (a point we illustrate in Section VII).

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 hint at ways to improve existing measures.

First, using the weights on life domains elicited in our survey, we broadly replicate past findings from the literature, discussed in Section II, 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, forward-looking, or flow utility. The three commonly used SWB measures resemble flow more than lifetime or forward-looking utility, but they seem to also put weight on the past and future. We also find that the five new 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, this finding again suggests that by rewording SWB questions, researchers may be able to nudge respondents in the direction of a desired utility notion (in this case, a desired time-horizon profile). One finding we did not anticipate is that, among the eight SWB questions we study, a new question we authored about "personal well-being" evokes the most flow-like time-horizon profile.

Third, with regard to social circles, on average across respondents, the eight SWB measures we study look more similar to each other than they do with regard to time horizons. Respondents report putting the highest weight on themselves and second highest on their immediate family, with wider social circles receiving less weight. Thus, our results suggest that the 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. For example, changing the wording of an SWB question from "personal" to "family" well-being yields dramatically more family-centered weight profiles. We also again have some unexpected findings, such as that the standard "ladder" SWB question yields more self-centered weights than other standard SWB questions (life-satisfaction and happiness).

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 and family-centered utility more for women and the unemployed; men and the employed report putting higher relative weight on broader time horizons that cover their entire lives and on broader social circles that go far beyond their immediate family.

In Section VII we summarize results from an analysis (fully reported in Web Appendix Section 2) that illustrates how, regardless of whether SWB approximates some utility notion on average, the heterogeneity in time-horizon and social-circle weights across sociodemographic groups can create an additional confound for scientific and policy conclusions. We focus on a common application of SWB data: regressions of SWB on respondent characteristics, often called "happiness regressions." We first estimate a typical happiness regression. We then re-estimate it with additional control variables, which are constructed from our survey-elicited weights to quantify respondent-level differences in weight profiles. We find that some sociodemographic coefficients are sensitive to the inclusion of self-versus-others-profile controls (but not time-profile controls, possibly due to measurement error in these controls). Our results imply that conclusions of cost-benefit analyses that rely on SWB coefficient magnitudes—including, e.g., pricing unemployment in dollars, as mentioned above—can be partly driven by differences across respondents in the weights they use when answering the SWB question.

Section VIII concludes, briefly describes other introspective questions that we explored in our survey, and discusses broader implications of our findings. While our findings suggest

caution in assuming that the SWB questions we study 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 eight SWB questions we study, a newly phrased Personal Well-Being question—"On a scale from 0 to 10, how would you rate your overall personal well-being?"—comes closest to eliciting self-centered flow utility and thus may be a useful point 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.

I. Theoretical Framework

Our theoretical framework clarifies different possibilities assumed in the SWB literature regarding what preference information might be captured by an SWB question. This framework underlies our discussion of the literature in Section II and our survey design and analysis in the rest of the paper.

I.A. Preferences

As we discuss below, papers in the SWB literature often assume that, rather than capturing the full preferences that are revealed by choices, 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

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

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 of person $k \neq 0$ in the

individual's (overall) flow utility, and $u_k(\boldsymbol{c}_{t,k})$ has standard properties and is defined up to a positive affine transformation. The consumption vectors consist of D domains of life: $\boldsymbol{c}_{t,k} = (c_{t,k,1}, c_{t,k,2}, \ldots, c_{t,k,D})'$.[2]

For convenience, we normalize the discount factors so that $\delta_{t,t} = 1$. In most applications, economists assume exponential discounting ($\delta_{t,\tau} \equiv \delta^{\tau-t}$). 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—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 strong tradition in welfare economics of using lifetime utility. Period t's *flow utility* is $u(\boldsymbol{c}_t) \equiv \sum_{k=0}^{K} \lambda_k u_k(\boldsymbol{c}_{t,k})$ where $\boldsymbol{c}_t = (\boldsymbol{c}'_{t,0}, \boldsymbol{c}'_{t,1}, \ldots, \boldsymbol{c}'_{t,K})'$.

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] 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 IV when discussing how our survey handles status concerns.

I.B. 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:

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

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 $\widetilde{U}_t$ as the *SWB function*. The resulting value of $\widetilde{U}_t$ 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 (discussed in Section II) 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 $\tau \neq t$.

For completeness, we note that 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. Although it does not correspond to any standard utility notion, and therefore is not a focus of this paper, backward-looking utility might be natural for survey respondents asked to reflect upon their lives.[3]

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-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 $\widetilde{U}_t$ 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 $\widetilde{U}_t$ 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 $\widetilde{U}_t$ can be interpreted as *any* monotonic transformation of a standard utility notion U_t.

[3] Some SWB questions (which we do not study) are explicitly backward-looking. Examples include questions about happiness and other emotions yesterday or in the past week (as in many surveys, including 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). The use of such SWB questions is not motivated by some backward-looking utility notion, but rather by their potential to be convenient measures of just-experienced flow utility, or of the integral of flow utility over a recent period of time. 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).

I.C. Identification Strategy

In the empirical part of this paper, our survey aims to shed light on whether, and to what extent, responses to an SWB question satisfy some of the necessary conditions for the above assumptions to hold. As we detail there, we ask respondents how they weighted different life domains, time horizons, and social circles when they answered an SWB question. The survey can therefore inform us about the components $\tilde{u}_k(\cdot)$, $\tilde{\delta}_{t,\tau}$, and $\tilde{\lambda}_k$ of the SWB function $\tilde{U}_t$ of each SWB question we study. However, the survey provides no information about the components $u_k(\cdot)$, $\delta_{t,\tau}$, and λ_k of the utility function U_t. Thus, we cannot *directly* test $\tilde{\delta}_{t,\tau} = \delta_{t,\tau}$, $\tilde{\lambda}_k = \lambda_k$, or $\tilde{u}_k(\cdot)$ equals a positive affine transformation of $u_k(\cdot)$.

Our survey is informative, however, 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. This allows us to draw certain conclusions about the assumptions above under certain conditions, as we describe in more detail in Sections IV–VI. For example, finding $\tilde{\delta}_{t,\tau} \neq 0$ for $\tau < t$ 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—as we detail in Section III—respondents are randomly assigned into one of eight groups, each answering a different SWB question, preferences are on average the same across the groups; any differences in weights found across the eight SWB questions must therefore reflect differences across SWB functions. This point underlies our falsification tests, which look for sensible changes in weights in response to changes in 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.

In the next section, we discuss existing work in the SWB literature in light of our theoretical framework before turning to our survey and results in the remainder of the paper.

II. Assumptions About Utility in the SWB Literature

Specific applications in the SWB literature often rely on many strong assumptions, including theoretical (e.g., homogeneous preferences), measurement-related (e.g., homogeneous response-scale use), and econometric assumptions (e.g., no omitted variables). We are skeptical

of many of these assumptions, but here we take them for granted. Instead, in this section, we focus on the assumptions made regarding the relationship between $\widetilde{U}_t$ and U_t. In Section II.A, we discuss prior literature that tests whether $\widetilde{U}_t$ is a monotonic transformation of U_t. In Sections II.B and II.C, respectively, we discuss assumptions made about the time horizon and social circles captured by $\widetilde{U}_t$ (possibly narrower than those captured by U_t).

II.A. SWB = Preferences

According to some leading philosophical conceptions of well-being (e.g., Railton, 1986), preferences are identified with the choices a person would make after deliberation when well informed about the consequences of her options. In our framework, SWB would be a measure of such preferences—i.e., the utility function that is maximized by well informed, deliberated choices—if the SWB responses capture (i) the same function of the different life domains as utility, (ii) a utility notion whose forward-looking component coincides with forward-looking utility (any backward-looking component is irrelevant for choice, under the assumption that preferences are time-separable), and (iii) other-regarding preferences.

One approach to evaluating the extent to which SWB is a good measure of preferences 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 $\widetilde{U}_t$ is not a positive monotonic transformation of U_t and, in particular, $\widetilde{U}_t$ 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). 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 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 this literature 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.[5] 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.

II.B. Time Horizon

[4] Here are examples: the MRS 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] 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.

Few papers that apply SWB data explicitly discuss which intertemporal preference information is captured by the SWB measure. Exceptions include Gruber and Mullainathan (2005) and Blanchflower and Oswald (2004), both of whom analyze the General Social Survey (GSS) happiness question ("Taken all together, how would you say things are these days—would you say that you are very happy, pretty happy, or not too happy?"). In their 2002 working paper, Gruber and Mullainathan (pp. 24, 28–29) argue that their evidence is most consistent with forward-looking utility. In contrast, Blanchflower and Oswald (p. 1362) state that the *same* question is "more naturally interpreted as a flow rather than a stock." Another exception is Alesina, Di Tella, and MacCulloch (2004), who also analyze this same GSS happiness question, as well as a life satisfaction question from the Euro-barometer Survey. Referencing Kahneman, Wakker, and Sarin's (1997) concept of "experienced utility" as a flow of affective experiences[6], Alesina et al. appear to interpret their SWB questions as capturing flow utility: "Our paper, and we believe much of the happiness literature, can be understood as an application of experienced utility, a concept that emphasizes the pleasures derived from consumption" (their footnote 7). Yet later, after finding that greater national inequality reduces individuals' SWB, they argue that a plausible mechanism is beliefs about how national inequality affect one's own future prospects—a mechanism that is more naturally interpreted under the assumption that SWB captures forward-looking or lifetime utility.

The other two papers we are aware of that make explicit statements about their time-horizon assumptions are Finkelstein, Luttmer, and Notowidigdo (2013) and Aghion, Akcigit, Deaton, and Roulet (2016). Finkelstein et al. (2013) study the Health and Retirement Study happiness question ("Much of the time during the past week I was happy. Would you say yes or no?") and write (p. 234): "As is standard in the happiness literature, we interpret the happiness question as a proxy for von Neumann–Morgenstern (flow) utility." Aghion et al. (2016) treat the Cantril (1966) ladder question from the Gallup Healthways Well-Being Index and the life satisfaction question from the Behavioral Risk Factor and Surveillance System (BRFSS) survey

[6] Kahneman, Wakker, and Sarin themselves are not only explicit about assuming that their SWB question is measuring "instant utility" but they also use SWB questions that ask a respondent to answer about the immediate present—an example of consistency between assumption and SWB question that we argue should be followed more generally. However, the specific SWB questions they study, such as "rate the current intensity of pain" prompted every 60 seconds (Redelmeier and Kahneman, 1996), have not been widely adopted in the SWB literature and so are not a focus of this paper.

as measures of forward-looking utility: "Life satisfaction is captured by the expected discounted valuation of an individual's future earnings" (p. 3870).

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

First, it is common in economic applications to calculate the money valuation of a good. Specifically, researchers run a regression of current SWB on current income y_t and current consumption of a "good" x_t (in our framework, this would be one component of the vector of life domains $\boldsymbol{c}_t$), and calculate the money value of the good as the ratio of coefficients. The appropriate interpretation of the resulting value depends on what preference information is captured by the SWB data. For example, if SWB measures flow utility $\widetilde{U}_t = u(\boldsymbol{c}_t) \equiv u_t$, then this coefficient ratio is $ratio_u \equiv \frac{\mathrm{Cov}(u_t,\breve{x}_t)/\mathrm{Var}(\breve{x}_t)}{\mathrm{Cov}(u_t,\breve{y}_t)/\mathrm{Var}(\breve{y}_t)}$, where $\breve{x}_t$ is the residual from a regression of x_t on all covariates (including y_t) and $\breve{y}_t$ is the residual from a regression of y_t on all covariates (including x_t). If SWB measures a broader notion of utility $\widetilde{U}_t = U_t$, such as lifetime or forward-looking utility, then the coefficient ratio is $ratio_U \equiv \frac{\mathrm{Cov}(U_t,\breve{x}_t)/\mathrm{Var}(\breve{x}_t)}{\mathrm{Cov}(U_t,\breve{y}_t)/\mathrm{Var}(\breve{y}_t)}$. These ratios are equal if and only if $\frac{\mathrm{Cov}(U_t-u_t,\breve{x}_t)}{\mathrm{Cov}(u_t,\breve{x}_t)} = \frac{\mathrm{Cov}(U_t-u_t,\breve{y}_t)}{\mathrm{Cov}(u_t,\breve{y}_t)}$. In words, current $\breve{x}_t$ and $\breve{y}_t$ must have the same ratio of covariance with non-current utility, $U_t - u_t$, to covariance with current utility, u_t. This condition would hold if, for example, the regression coefficients were identified by exogenous shocks in $\breve{x}_t$ and $\breve{y}_t$ that affect flow utility only in period *t*. However, in practice the condition is unlikely to hold because in nearly all applications, the regression coefficients are identified by cross-sectional variation, and $\breve{x}_t$ and $\breve{y}_t$ typically have different time profiles of covariance with flow utility.

We illustrate with the example of calculating the money valuation of unemployment status, a common application in the SWB literature (e.g., Blanchflower and Oswald, 2004). Here, x_t is employment status. Consider three assumptions that dramatically simplify a more complicated reality but help make the point clear: (a) unemployment conditional on the covariates occurs randomly, lasts one period, and only affects current flow utility; (b) the effect of current income conditional on the covariates, $\breve{y}_t$, on current flow utility is identical across periods; and (c) cross-sectional variation in current income y_t fully reflects cross-sectional

variation in permanent income. Under assumption (a), $\text{Cov}(U_t, \breve{x}_t) = \text{Cov}(u_t, \breve{x}_t)$, so the numerators of $ratio_u$ and $ratio_U$ are equal. However, under assumptions (b) and (c), the denominators of $ratio_u$ and $ratio_U$ differ substantially; for example, if U_t is forward-looking utility and $t < T$, then $\text{Cov}(U_t, \breve{y}_t) = \sum_{\tau=t}^{T} \delta_{t,\tau} \text{Cov}(u_\tau, \breve{y}_\tau) \gg \text{Cov}(u_t, \breve{y}_t)$. In words, $ratio_u$ is an estimator of the "dollar cost of unemployment" in units of current income—that is, a one-time cost—whereas ratio_U is an estimator of this cost in units of permanent income—that is, an annually recurring cost. Thus, if the SWB measure is assumed to capture forward-looking utility instead of flow utility, then the (same) numerical estimate from the SWB regression implies a cost of unemployment many times larger!

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

Second, another common application of SWB data, especially in the psychology literature, is to infer that people are making mistakes when a deviation from usual behavior is found to increase SWB. For example, Dunn, Aknin, and Norton (2008) find that experimental participants randomly assigned to spend money on someone else rather than on themselves were happier when surveyed later the same day. This and related findings, together with survey evidence that people expect spending on themselves to make them happier than spending on others, lead Dunn et al. to conclude that "policy interventions that promote prosocial spending…may be worthwhile." However, the inference that people are making a mistake is only warranted if the forward-looking component of the SWB function represents forward-looking utility. Otherwise, apparent "mistakes" could instead reflect optimal intertemporal tradeoffs. Indeed, consistent with this possibility, Falk and Graeber (2020) find that experimental participants randomly assigned to donate money to charity rather than receive money themselves were *less* happy four weeks later (possibly due to the foregone consumption on self), despite being happier at the end of the lab session.

Finally, there is a growing literature on how SWB varies with age. Much of the evidence points to a U-shape, with SWB reaching a nadir in middle age, but there is no consensus on the

reason for this pattern (e.g., Blanchflower and Oswald, 2008; Blanchflower, 2020). While people might change how they use the SWB response scale with age (as suggested by Stone, Schneider, Junghaenel, and Broderick, 2019), few analyses account for this possibility; therefore, consistent with the literature, our discussion here assumes that, with age, the SWB function $\widetilde{U}_t$ remains the same monotonic transformation of the utility notion U_t that a particular paper focuses on.

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

Moreover, the interpretation of the SWB measure is relevant for what mechanisms may explain the U-shape. If it is lifetime utility, then the change in SWB from age t to $t+1$ is $\widetilde{U}_{t+1} - \widetilde{U}_t = \sum_{\tau=0}^{T}\left[E_{t+1}\delta_{t+1,\tau}u(\boldsymbol{c}_\tau) - E_t\delta_{t,\tau}u(\boldsymbol{c}_\tau)\right]$. In that case, changes in SWB with age are due to differences between $\delta_{t+1,\tau}$ and $\delta_{t,\tau}$ and to unanticipated shocks to flow utility.[7] In contrast, if SWB measures flow utility, then variation with age is unrelated to both discount factors and beliefs: $\widetilde{U}_{t+1} - \widetilde{U}_t = u(\boldsymbol{c}_{t+1}) - u(\boldsymbol{c}_t)$, implying that the quantity or quality of consumption in at least some domains of life increases with age.

II.C. Social Circles

Just as with intertemporal preferences, few SWB applications discuss which other-regarding preference information is captured by the SWB measure. While we are not aware of papers that are explicit and unambiguous, Ludwig et al.'s (2012) discussion of the GSS

[7] Schwandt (2016) studies unanticipated shocks directly, with data on predicted and subsequently realized SWB. Note that for lifetime utility, it may be natural to discard our normalization $\delta_{t,t} = 1$ and instead assume that the discount factor that applies to a particular age does not depend on current age: $\delta_{t,\tau} = \delta_{t',\tau} \equiv \delta_\tau$ for all t, t', τ (a special case is equal weighting of each period τ). In that case, $\widetilde{U}_{t+1} - \widetilde{U}_t = \sum_{\tau=0}^{T}\delta_\tau[E_{t+1}u(\boldsymbol{c}_\tau) - E_t u(\boldsymbol{c}_\tau)]$, so changes in SWB with age are *entirely* due to unanticipated shocks to flow utility.

happiness question suggests that they treat responses as a measure of family well-being: "Another reason we focus on adults is because more is known about measuring SWB of adults than youth … [SWB] was added to the long-term survey to be one of the key summary measures of the net impacts on families …." (p. 1507). Easterlin (1995) studies this same GSS happiness question and a variety of international life satisfaction data sources and offers an interpretation that is ambiguous: "Formally, this model corresponds to a model of interdependent preferences in which each individual's utility or subjective well-being varies directly with his or her own income and inversely with the average income of others" (p. 36). Easterlin's mention of "interdependent preferences" sounds like other-regarding utility, but the rest of the sentence sounds more like SWB is capturing self-centered utility that depends on relative income, as in Frank's (1985) model of status concerns. Similarly, Alesina, Di Tella, and MacCulloch (2004) treat this same GSS happiness question, as well as a life satisfaction question from the Eurobarometer Survey, as measuring some notion of utility but do not specify if the utility notion is other-regarding or self-regarding (with respondents concerned about how inequality affects their own future prospects): "In this paper, we explore whether and why inequality negatively affects individual utility even after controlling for individual income. We measure 'utility' in terms of survey answers about 'happiness'" (p. 2010).

Again, as with time horizon, the social-circle interpretation of the SWB question often matters in SWB applications even when it is not discussed. Here are four examples. First, consider the finding mentioned above that giving to others increases short-run happiness (e.g., Dunn, Aknin, and Norton, 2008). How much this can be attributed to altruistic preferences, as opposed to self-signaling (e.g., Bénabou and Tirole, 2006) or warm glow (Andreoni, 1989), depends on the extent to which the SWB data capture other-regarding preference information. Second, a number of papers compare SWB between men and women (e.g., Stevenson and Wolfers, 2009). If SWB is capturing family well-being rather than self-centered utility, then for individuals who love their opposite-sex family members, the SWB difference between men and women may understate the self-centered well-being difference. Moreover, if men interpret the SWB question as applying to a broader or narrower social circle than women do, and especially if such interpretational differences vary over time, then these interpretational differences confound conclusions from SWB data about self-centered well-being differences between men and women. Third, consider papers that compare SWB between people with and without

children (e.g., Deaton and Stone, 2014). Such comparisons are usually construed in terms of self-centered utility; they are much harder to understand if SWB captures family well-being and hence includes a concern for the children. Finally, consider papers and policy reports that compare mean SWB across countries (e.g., the United Nation's annual World Happiness Report; WHR, 2021). When such comparisons are interpreted as national well-being rankings, the use of mean SWB amounts to adopting a social welfare function that is utilitarian. Moreover, since the arguments of a social welfare function are self-centered utilities (for a recent review, see Adler, 2019), such applications implicitly assume that SWB is measuring self-centered utility.

Summarizing the first part of the paper, the SWB literature generally assumes, implicitly or explicitly, that SWB data measure some utility notion—often narrower than the utility corresponding to choice—but makes a variety of different assumptions regarding *which* utility notion it is. To shed some initial empirical light on how respondents *actually* interpret different SWB questions, we now turn to the second part of the paper: the design and analysis of our new survey.

III. Survey Design

In our 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.[8] 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 Section 5 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

[8] 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; Junghaenel 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.

respondents, and their answers to the SWB question. We defer describing the follow-up questions to subsequent sections, where we discuss the design of these questions, their links to the theory from Section I, and the empirical findings from those questions.

III.A. SWB question

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 Office for National Statistics Integrated Household Survey. 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 (as in much of the work cited earlier).

The remaining five questions are new questions that, to the best of our knowledge, have not been previously used in applied work. We include them in the survey with the general aim of exploring the potential of new SWB questions to "do better" for our purposes than currently used questions. By "do better" we mean that they may (1) more closely 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 the kinds of applied work discussed in Sections II.B and II.C.

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, since the specific dimensions elicited by these "eudaimonic" SWB questions may not be fully captured in standard evaluative SWB questions (e.g., Ryff, 1989), there have been proposals to include them 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 main 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.

III.B. 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 Section 1 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 (5^{th}- and 95^{th} percentiles were 6 and 57 minutes).

III.C. Responses to SWB question

Figure 1 reports histograms summarizing responses to the main SWB questions, by SWB question and (at the bottom right) pooled. The overall median response is 8 on a 0–10 scale (5^{th} and 95^{th} percentiles are 2 and 10). Looking question by question, the median response is 8 in all but the Ladder (median = 6) and the Meaning & Value (9) questions. For each question, 5^{th} and 95^{th} percentiles are 1–4 and 10.[9] The median time to answer the SWB question was 12.6 seconds (5^{th} and 95^{th} percentiles were 5.4 and 57.6 seconds).[10]

IV. Weights on Life Domains and Tests of Introspective Methodology

[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 new question that 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 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.

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 life[11] in random order, and a sixteenth "Other (please specify)" entry (always at the bottom), each with a slider labeled from "Not at All" to "A lot."[12] Due to relatively few responses to the "Other" slider on this and other survey screens, here and in the rest of the paper we do not include the "Other" option in the analysis.

IV.A. 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

[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' concerns about social status are likely to be reflected in their responses to our social-circles questions (described in Section VI 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 $\boldsymbol{c}_{t,0} - \boldsymbol{c}_{t,k}$. We do not pursue this extension because the exposition of the model in Section I 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.

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 in our survey. The domains "Income & financial security," "Family life & relationships," "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 II.A 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 II.A finds 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 II.B above, those correlations are between $\widetilde{U}_t$ MRSs and U_t MRSs.[13] To investigate the information captured by our slider-based domain weights, we similarly compare

[13] Formalizing these 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 $\widetilde{U}_t$ 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 $\widetilde{U}_t$ 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: $\frac{\partial U_t}{\partial c_{t,0,d}} \Big/ \frac{\partial U_t}{\partial c_{t,0,d'}} = \frac{\partial u_0}{\partial c_{t,0,d}} \Big/ \frac{\partial u_0}{\partial c_{t,0,d'}}$ and $\frac{\partial \widetilde{U}_t}{\partial c_{t,0,d}} \Big/ \frac{\partial \widetilde{U}_t}{\partial c_{t,0,d'}} = \frac{\partial \tilde{u}_0}{\partial c_{t,0,d}} \Big/ \frac{\partial \tilde{u}_0}{\partial c_{t,0,d'}}$. 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 $\widetilde{U}_t$ 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 II.A above, the MRSs across life domains for U_t and $\widetilde{U}_t$ 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.

them with MRS estimates for U_t from Benjamin, Heffetz, Kimball, and Szembrot (2014, henceforth BHKS)—mentioned in III.A above—who use 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 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 $\widetilde{U}_t$, 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 $\widetilde{U}_t$, and that similarly, our slider-based time-horizon and social-circles weights provide substantial information about $\tilde{\delta}_{t,\tau}$ and $\tilde{\lambda}_k$.

IV.B. 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 III.A. 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 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.")

[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).

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.

IV.C. 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 questions[15] 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 cited above.

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 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

[15] Web Appendix Figures I–III reproduce 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 III.A 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.

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.

IV.D. 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 IV.A 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 IV.B above, and throughout the paper).

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).

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

V. Weights on Time Horizons

To investigate the time period over which respondents evaluated their situation when answering the SWB question, we ask 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, $u(\boldsymbol{c}_\tau)$, 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 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.

---

[16] Here and in our theoretical framework in Section I, 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 Junghaenel et al. (2018).

[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 be less able to accurately respond to such sliders (and, more generally, would find them cumbersome and confusing). (Moreover, from a theoretical point of view, avoiding overlap is not necessary. For example, any combination of relative weights for "This moment" and "Today, excluding this moment" can be replicated as a combination of "This moment" and "Today" (including this moment).)

V.A. General patterns

Figure 4 provides three (selected) example individual-level responses; all 3040 individual responses are reported in the Individual Responses Web Appendix. Respondent #2559 pushed the "Right this moment" and "Today" sliders to the extreme right and kept all other sliders at the extreme left (we later coded these as 100 and 0, respectively). We view such a response pattern as consistent with flow utility with period-*t* duration around a single day. Alternatively, it is also consistent with lifetime or forward-looking utility with extremely myopic preferences, but (as we argue below) we think that is unlikely. Respondent #2162, on the other hand, pushed the "Entire life including your expectations for the future" slider to 100, and left all other sliders at 0, consistent with lifetime utility. Of course, one should not expect such clean patterns in mean responses or, indeed, for most single respondents. For example, on neither extreme, respondent #1175 reports a more complex combination of weights that suggests more focus on the present and the past than on the future and that does not naturally fit into one of the main utility notions in economic applications.

Aggregating across all respondents, 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 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 (the extreme myopia mentioned above).

V.B. 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 Section 3). 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.

V.C. 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.

VI. Weights on Social Circles

To explore *whose* well-being respondents considered in answering the SWB question, our respondents were first asked:

> 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.

VI.A. 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 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.

VI.B. Comparing across SWB questions

As seen in the rightmost graph, the comparison between Personal Well-Being ("your overall personal well-being") and Family Well-Being ("the overall well-being of you and your family") again suggests that respondents react sensibly to the wording of the SWB questions.

[18] In the remainder of Section VI, 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 Section VII 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). Web Appendix Section 2 provides full details, and shows that results remain very similar across alternative specifications, including specifications that include all respondents by imputing values for the missing follow-up-screen responses.

The two questions are nearly identical except for the mention of family. Consistent with this one difference, we find that they virtually coincide on all sliders other than "Yourself" and "Immediate Family," while they differ dramatically on these two: the respective weights are 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, $\tilde{\lambda}_k = 1$ 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.

The center graph shows that 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. Their higher weight on larger social circles than other questions is consistent with other-regarding preferences: $\tilde{\lambda}_k > 0$ for all k.

Finally, for Meaning & Value, and to a lesser extent for Options & Possibilities, respondents assign higher weights on 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 IV.B above that these SWB questions are associated with higher weights on the domains "Volunteering, activism" and "Family life & relationships."

VI.C. 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, with the possible exception of the age groups, which may differ on the ranking of the wider social circles (beyond self and immediate family) relative to each other. However, we note that relative to men, younger, and employed respondents, women, older, and unemployed respondents have higher or equal mean weights on one's self and immediate family but lower mean weights on wider social circles. We also see corresponding ordinal differences at the *individual* level, with a smaller fraction of men, younger, and employed respondents ranking one's self and immediate family higher. We analyze

the (ordinal) individual-level metrics of closeness to flow and family-centered notions in the next section.

VII. How are cross-group SWB comparisons affected by differences in weights?

Much of the SWB literature in economics focuses on cross-group comparisons of responses to SWB questions. Such comparisons assume that SWB responses capture the same utility notion across the groups. However, in Sections V.C and VI.C, we record evidence of cross-group weighting differences for time horizons and social circles, which casts doubt on this assumption, thus creating a confound for interpreting findings of cross-group SWB differences.

In this section, we summarize results from Web Appendix Section 2, where we explore how these weighting differences may affect conclusions about cross-group SWB comparisons. In the earlier sections, we studied the entire profile of slider responses but examined only univariate sociodemographic splits one at a time (age, sex, income, and employment status) averaged across respondents. To facilitate comparisons with the literature, in this section we switch to a multivariate regression framework for the sociodemographics and summarize the slider responses with respondent-level summary indexes, as described below.

Conceptually, our approach has three steps (the same steps as in a mediation analysis, albeit with a somewhat different interpretation): (a) run a standard regression of respondents' SWB responses on a full set of available sociodemographics in our survey data; (b) re-run the regression but additionally control for each respondent's weight profile on time horizon and/or social circle; then (c) examine how the coefficients on the sociodemographics are affected by the additional controls. To increase statistical power and reduce multiple hypothesis testing, we pool data from all the SWB questions, and we implement step (b) using only two control variables: one summarizing time-horizon weights and one summarizing social-circle weights.

In the Web Appendix, we examine several alternative definitions of these weight-profile control variables (both ordinal and cardinal) to ensure robustness of our findings. As we show there, under some linearity assumptions and assuming that the weight-profile controls have no measurement error, the regressions with both controls tell us what the coefficients on the sociodemographic characteristics would be if SWB were understood by everyone equally to be fully flow utility and fully family-centered utility.

In practice, our weight-profile control variables have measurement error. If the measurement error is classical and uncorrelated with the sociodemographics, then the differences in coefficients on the sociodemographics in step (c) are lower bounds on what the changes would be from controlling for non-noisy measures.[19]

Our results from step (a)—a "happiness regression" of SWB on sociodemographics—broadly mirror those that have been found in the literature: SWB is higher among respondents who have higher income, are more educated, more religious, older, and married; and lower among the unemployed. We also find that in our data, SWB is higher among women.

In steps (b) and (c), we find that controlling for our time-horizon weight profile has no effect on regression coefficients, but that controlling for our social-circle weight profile results in several meaningful coefficient changes. In our benchmark specification (Web Appendix Table III), the coefficients on non-white, religious, and unemployed, for example, increase in magnitude by 57, 13, and 7 percent, while those on old and female shrink by 15 and 10 percent, respectively. Such large changes would have a substantial impact in applications that rely on coefficient magnitudes, such as efforts to "price" the costs of unemployment in terms of the decrease in income associated with the same decrease in SWB (e.g., Clark and Oswald, 2002). Our results also suggest that the increase in SWB at older ages (associated with the U-shape of SWB with age) is partly driven by social-circle weight profiles becoming increasingly family-centered with age (this finding of ours is consistent with socioemotional selectivity theory, which posits that as people age, they prioritize close relationships and obtain more satisfaction from them; for a review, see Löckenhoff and Carstensen, 2004). Our findings thus serve as a caution that scientific and policy conclusions that depend on coefficient magnitudes may sometimes be driven by cross-group differences in the weights regarding whom the SWB question applies to.

VIII. 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

[19] For the case of a single control variable measured with error, the claim follows directly from known results (Garber and Klepper, 1980; for a direct proof, see https://blog.supplysideliberal.com/post/2019/10/10/adding-a-variable-measured-with-error-to-a-regression-only-partially-controls-for-that-variable). Therefore, the change in sociodemographics' coefficients when we control for one of the profile variables is a lower bound on what the change would be from a non-noisy measure of the variable. The same logic then applies iteratively when we additionally control for the other profile variable.

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 document that economists and other social scientists make a variety of assumptions, sometimes inconsistent with each other, about what utility notions are captured by SWB questions, and we argue that the conclusions that can be drawn from many of the applications of SWB data hinge on which assumption is made. Empirically, we 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, none of the SWB measures we studied, including both those based on standard happiness, life satisfaction, and ladder questions, and new ones that we devised, have the time profile of flow utility, forward-looking utility, or lifetime utility. Second, none of the measures corresponds to self-centered utility but 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, first, substantial differences in weights about time profiles and social circles. Second, we find that in some cases of comparing SWB across sociodemographic groups—such as younger vs. older—controlling for differences in social-circle weights across respondents has a substantial impact on the estimated coefficient, even with our imperfect controls for introspection profiles.

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 long-term average, aggregating these disparate data into a normative measure of welfare is challenging.

Two of our findings point directly to readily applicable practical advice for researchers. First, we find that small adjustments to the wording of SWB questions are effective in shifting respondents' weights in the expected direction. Our advice based on this finding depends on the researchers' latitude to shape the survey data they analyze. To users of existing SWB data, we caution against 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.[20] Among the eight SWB questions investigated in this paper, our respondents report that 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 self-centered flow utility, comes the closest to doing so. This short and simple question may be a promising point 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 different 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.

Second, we find that when we control for differences across respondents in their time-horizon and social-circle weights, the coefficients from a regression of SWB on sociodemographics change, in some cases substantially. Moreover, since 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).[21] We caution,

[20] For example, in a recent data-collection effort, Benjamin, Cooper, Heffetz, and Kimball (2019) include versions of the Personal Well-Being, Family Well-Being, and many other questions that, in addition to explicitly varying the relevant social circle, also explicitly state the relevant timeframe: "Thinking about the past year..." For another example, Allcott, Braghieri, Eichmeyer, and Gentzkow (2020) modify commonly used SWB questions to explicitly state "over the past four weeks," "over the last ten minutes," and "right now."

[21] 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).

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 (Benjamin, Cooper, Heffetz, and Kimball, 2020), such as sufficiently similar uses of the SWB response scales across the groups.

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 survey included additional introspective questions. We have not analyzed these data in detail, but for completeness, we mention them here and give some examples of questions and preliminary findings. We asked respondents, when they chose a particular number to respond to the SWB question, how much weight they put on thinking about how they *should* answer, thinking about their usual emotions and feelings these days, comparing their situation to other people, comparing to their own life in the past, comparing to their goals, and comparing to some absolute standard. We also asked follow-up questions about several of these possibilities. For the three standard SWB questions, we find that respondents report putting the highest weight on emotions/feelings, with a higher mean weight for Happiness than for Life Satisfaction or Ladder. Comparisons to one's goals and to one's past receive almost as much weight, and comparisons to an absolute standard generally receives the least weight. We view these preliminary findings as worthy of 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 well-being 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.

Clark, Andrew, and Andrew J. Oswald. 2002. "A simple statistical method for measuring how life events affect happiness." *International Journal of Epidemiology*, 31(6): 1139–1144.

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.

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.

Garber, Steven, and Steven Klepper. 1980. "Extending the Classical Normal Errors-in-Variables Model." *Econometrica*, 48(6): 1541–1546.

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.

Löckenhoff, Corinna E., and Laura L. Carstensen. 2004. "Socioemotional Selectivity Theory, Aging, and Health: The Increasingly Delicate Balance Between Regulating Emotions and Making Tough Choices." *Journal of Personality*, 72(6): 1395–1424.

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.

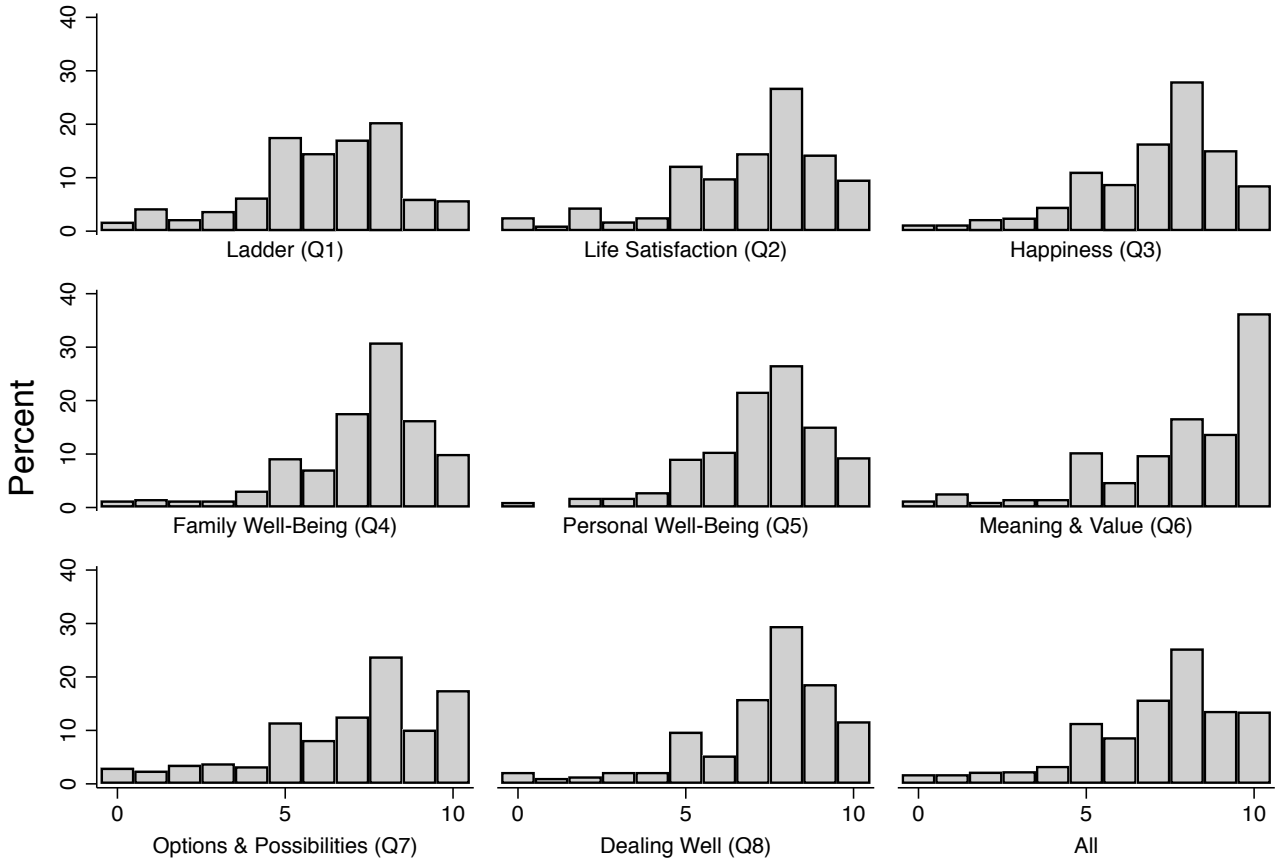
Stone, Arthur, Joseph Schwartz, Joan Broderick, and Angus Deaton. 2010. "A snapshot of the age distribution of psychological well-being in the United States." *Proceedings of the National Academy of Sciences,* 107(22): 9985–9990.

Szabó, Andrea, and Gergely Ujhelyi. 2017. "Choice and Happiness in South Africa." *Economics Letters*, 155: 28–30.

WHR. 2021. *The World Happiness Report*. https://worldhappiness.report

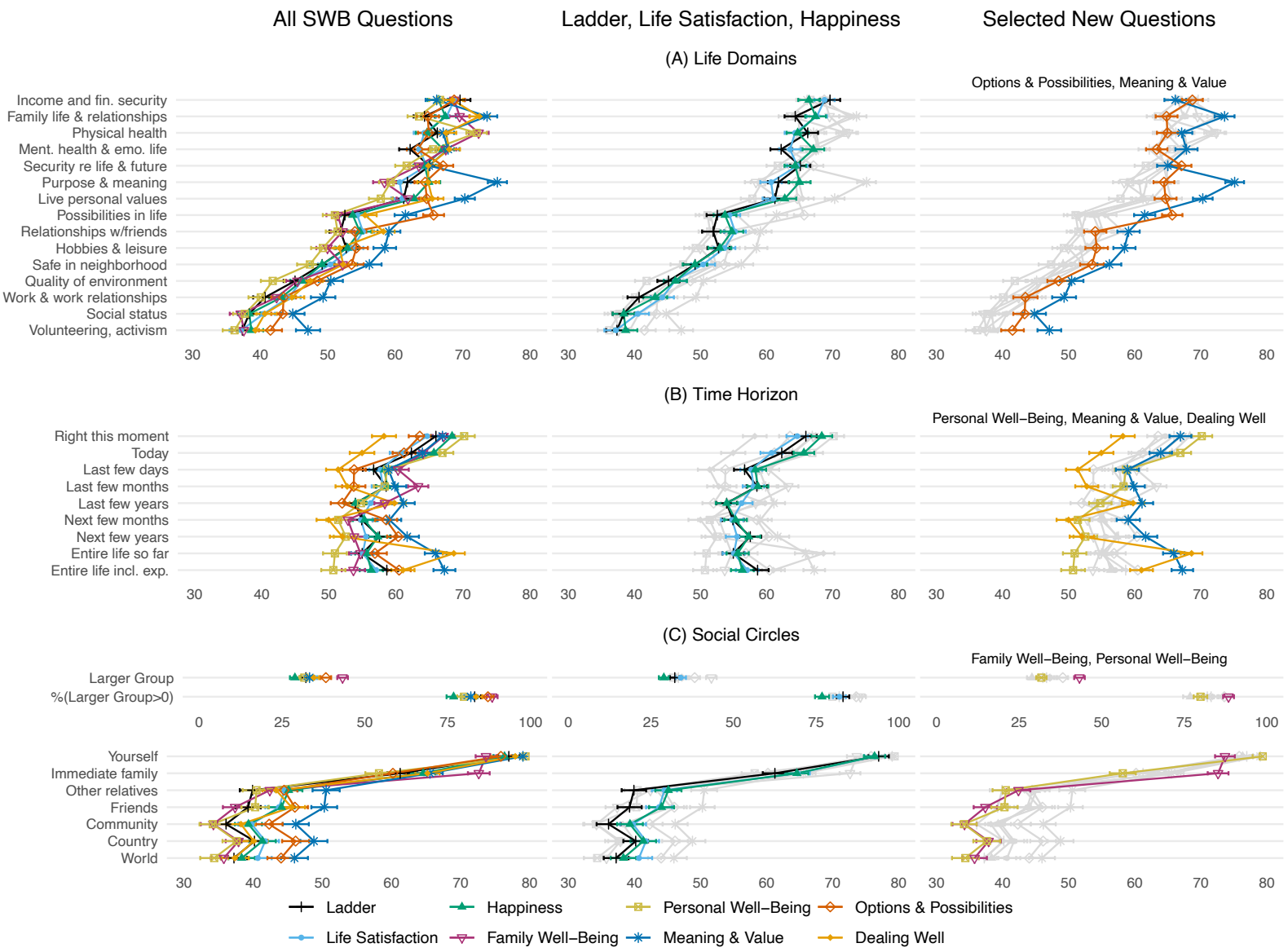
Zhou, Jiannan. 2020. "Survey Evidence on Habit Formation: Existence, Specification, and Implication." https://jiannanzhou.com/research/habit/Habit_Zhou_Paper.pdf

Figure 1: Response Histograms, by SWB Question



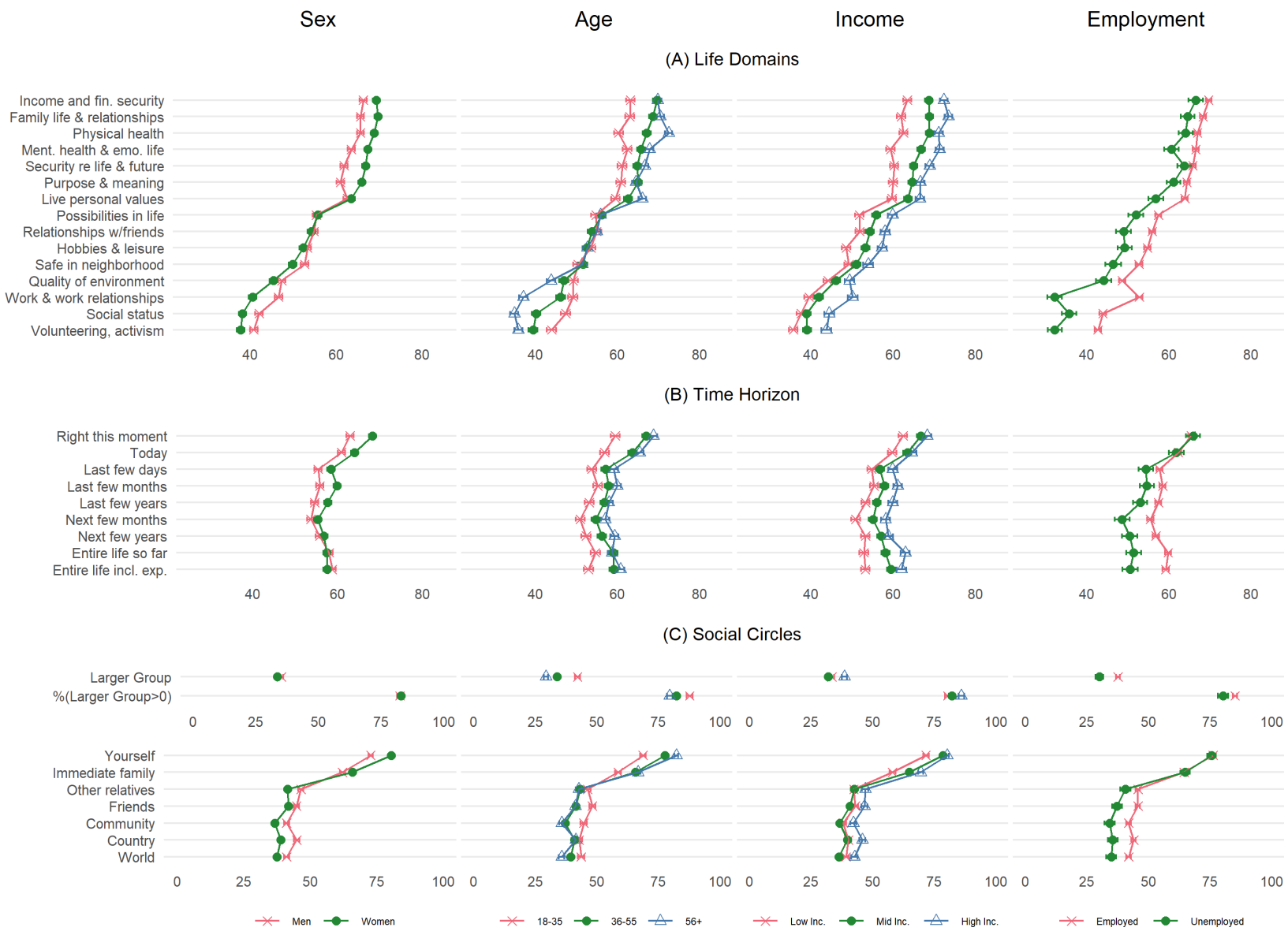
Notes: Based on 3,040 observations (All) and on 359–397 observations (each of Q1–Q8). Two respondents entered non-integers; these were rounded to the nearest integers.

Figure 2: Reported Weight Placed, by SWB Question



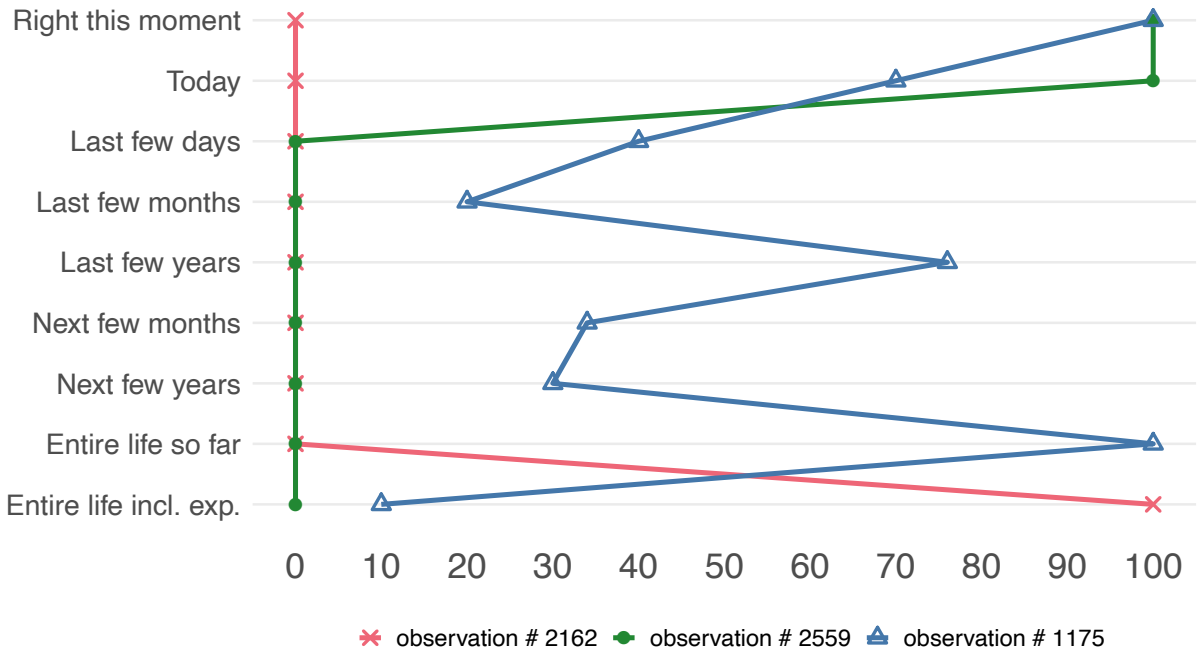
Notes: Based on 3,040 observations. Each row reports mean rating (0–100) by SWB question, other than "%(Larger Group > 0)" row, which reports percent of respondents who rated Larger Group above 0 (see text for details). "All SWB questions" column reports means/percent for all eight questions; "Ladder, 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



Notes: Based on 3,040 observations (fewer when demographic information is missing; see Appendix Table I). Each row reports mean rating (0–100) SWB question, 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: Example Reported Weights for the Time-Horizon Sliders (Ladder Question)



Notes: Slide responses for three (selected) survey respondents, all of whom had answered the Ladder SWB question.