

Beyond Happiness and Satisfaction: Toward Well-Being Indices Based on Stated Preference[†]

By DANIEL J. BENJAMIN, ORI HEFFETZ,
MILES S. KIMBALL, AND NICHOLE SZEMBROT*

This paper proposes foundations and a methodology for survey-based tracking of well-being. First, we develop a theory in which utility depends on “fundamental aspects” of well-being, measurable with surveys. Second, drawing from psychologists, philosophers, and economists, we compile a comprehensive list of such aspects. Third, we demonstrate our proposed method for estimating the aspects’ relative marginal utilities—a necessary input for constructing an individual-level well-being index—by asking ~4,600 US survey respondents to state their preference between pairs of aspect bundles. We estimate high relative marginal utilities for aspects related to family, health, security, values, freedom, happiness, and life satisfaction. (JEL C43, C83, D12, I31)

The cornerstone of neoclassical welfare economics is the principle of revealed preference, according to which the ultimate criterion for judging what makes a person better off is what she chooses, in a situation in which she is well informed

* Benjamin: Department of Economics, Cornell University, 480 Uris Hall, Ithaca, NY 14853, and NBER (e-mail: db468@cornell.edu); Heffetz: Samuel Curtis Johnson Graduate School of Management, Cornell University, 324 Sage Hall, Ithaca, NY 14853, and NBER (e-mail: oh33@cornell.edu); Kimball: Department of Economics, University of Michigan, 312 Lorch Hall, Ann Arbor, MI 48109, and NBER (e-mail: mkimball@umich.edu); Szembrot: Department of Economics, Cornell University, 455 Uris Hall, Ithaca, NY 14853 (e-mail: nes66@cornell.edu). We thank Gregory Besharov, Erzo F. P. Luttmer, Ted O’Donoghue, Andrew Oswald, Matthew Rabin, and Alex Rees-Jones for valuable early discussions and suggestions, and Matthew Adler, Aaron Bodoh-Creed, Alexander Cappelen, Viktoria Dalko, Angus Deaton, Koen Decancq, Marc Fleurbaey, Carol Graham, John Ifcher, Vithala Rao, Moses Shayo, and Bertil Tungodden for valuable comments. We thank three anonymous referees for suggestions that substantially improved the paper. For helpful feedback, we are grateful to participants at the Cornell Behavioral Economics Research Group and the Cornell Behavioral/Experimental Lab Meetings, and conference and seminar audiences at the WEAI annual meetings (2011, 2013), AEA annual meetings, IZA Workshop on Sources of Welfare and Well-Being, Michigan Social Psychology-Economics Workshop, Bay Area Behavioral and Experimental Economics Workshop, Technion’s Minerva Workshop on Decisions and Emotional Expressions, Samos AGT Summer School, NBER SI, Well-being in Contemporary Society Conference, Stanford Institute for Theoretical Economics, Paris Workshop on Well-Being and Preferences, New Directions in Welfare Economics III Conference, Aspen Conference on Economic Decisionmaking, Bank of Israel, Cornell, Hebrew University, London School of Economics, Norwegian School of Economics, Paris School of Economics, Tel Aviv University, UCLA, USC, Zurich, a University of Warwick short course on Cognitive Economics, and a Tel Aviv University mini-course on Happiness Measures in Economics. We thank Brice Cooke, Yeon Sik Cho, Samantha Cunningham, Thomas Davidson, Julia Goorin, Anthony Hawkins, Allen He, Julian Hsu, Sarina Kumar, Yupeng Li, Allison Nunez, Bora Park, Nathaniel Schorr, Atik Shah, Andrew Simon, Martha Widger, and especially Melissa Bickerman, Jaesun Lee, and Andrew Sung, for their research assistance. For financial support we are grateful to Cornell’s Institute for Social Science, Samuel Curtis Johnson Graduate School of Management, and NIH/NIA grants R01-AG040787, to the University of Michigan, and T32-AG00186 to the NBER. The authors declare that they received IRB approval from the relevant institutions and that they have no relevant or material financial interests that relate to the research described in this paper.

[†] Go to <http://dx.doi.org/10.1257/aer.104.9.2698> to visit the article page for additional materials and author disclosure statement(s).

about the consequences of her options. Yet for most policy decisions, a government cannot directly infer an individual's welfare from her choices over policies because individuals rarely make such choices.¹ Hence, in practice, economists often rely on revealed preference indirectly, evaluating policy options by how they affect indicators—most prominently, GDP—that can be viewed as summarizing, under some assumptions, a set of generally desired outcomes. But because GDP and other available indicators have known limitations as well-being measures, economists have been seeking additional indicators that go “beyond GDP” (for a recent survey, see Fleurbaey 2009). In this paper, we focus on developing one such indicator: an individual-level index that combines together different aspects of well-being that may be measured by survey questions.

As candidate measures of individuals' well-being, economists and psychologists have recently been investigating survey measures of “subjective well-being” (SWB); while we use this term to refer to any subjective assessment of some aspect of well-being, economists have primarily focused on questions about one's own happiness or life satisfaction. Because responses to such questions reflect a wide range of experiences, including those unrelated to market exchange (Diener and Seligman 2004; Kahneman and Krueger 2006), many researchers have advocated conducting nation-wide SWB surveys and using the responses to calculate indicators alongside GDP-like measures (e.g., Diener 2000; Diener 2006, signed by 50 researchers; Layard 2005; Stiglitz, Sen, and Fitoussi 2009).

Although these proposals are controversial among economists, policymakers have begun to embrace them. For example, starting in April 2011, the UK Office of National Statistics (ONS) began including the following SWB questions in its Integrated Household Survey, a survey that reaches 200,000 adults annually (ONS 2011):

Overall, how satisfied are you with your life nowadays?

Overall, how happy did you feel yesterday?

Overall, how anxious did you feel yesterday?

Overall, to what extent do you feel the things you do in your life are worthwhile?

According to Prime Minister David Cameron, “it's time we focused not just on GDP but on GWB—general wellbeing.”² Other governments around the world have expressed similar intentions to field SWB surveys and use the responses to guide policy.³

¹Holding a referendum on every issue would incur prohibitively high transaction costs. Moreover, for many issues, even a direct vote would not reveal preferences because voters lack full information about—and might systematically mispredict—the consequences of alternative policy options (see Gilbert 2006, for evidence on systematic misprediction of happiness).

²Allegra Stratton, “David Cameron Aims to Make Happiness the New GDP,” *The Guardian*, November 14, 2010, <http://www.guardian.co.uk/politics/2010/nov/14/david-cameron-wellbeing-inquiry> (accessed May 13, 2011).

³For example, in December 2011, the US National Academy of Sciences' Committee on National Statistics convened the first in a series of meetings of a “Panel on Measuring Subjective Well-Being in a Policy Relevant

Notwithstanding this recent enthusiasm, there are many open questions regarding the endeavor of tracking well-being with surveys. Among the most urgent practical questions are the following two. First, which SWB questions should governments ask? It is increasingly recognized that more than one question is likely to be needed because SWB is multidimensional (e.g., Ryff 1989; Kahneman and Deaton 2010) and because widely used SWB measures may not capture all factors that enter into preferences (Benjamin, Heffetz, Kimball, and Rees-Jones 2012, 2013). Current proposals for survey questions, however, rely on different experts' own readings of the SWB literature rather than on a systematic method.⁴ Second, how should responses to different questions be weighted relative to each other? Current proposals are virtually silent on relative weighting (in some cases purposefully so). But in practice, due to an apparently inevitable demand for summary indicators, ad hoc weights often end up being applied implicitly by users or explicitly in published indices (Micklewright 2001).

This paper has two overarching purposes. First, we propose a framework, grounded in a preference-based theory, for conceptualizing and discussing survey-based measurement of well-being. Second, we demonstrate a disciplined approach, anchored in revealed preference—albeit based on hypothetical choices—to applying our framework to the development of well-being surveys and indices. We emphasize that relative to the many decades of theoretical and practical work that underlies the present well-developed state of measures such as GDP, efforts to construct and apply survey-based well-being indicators are still in their infancy. Hence, we view this paper as primarily methodological, proposing an agenda for a new approach, and we view our specific contributions as first steps to be improved upon by future work.

In Section I, we present our theoretical framework. We assume that utility $u(\mathbf{w})$ depends on a vector \mathbf{w} of fundamental aspects of well-being, for example those that can be measured with survey questions similar to the four above. Any vector proportional to the vector of marginal utilities $D_{\mathbf{w}}u(\mathbf{w})$ can then be used as relative weights for combining the components of \mathbf{w} into an individual-level index that tracks small changes in well-being. For large changes in the aspects, the index can still be used to track changes in well-being but only provides a partial welfare ordering. While we do not make novel contributions regarding how to aggregate well-being indices across individuals, our framework could be used in conjunction with existing approaches to aggregation.

In Section II, we describe our attempt to identify the major components of \mathbf{w} . We compile a list—of 136 aspects of well-being—aimed at including all the main factors that have been proposed as important components of well-being in a sample of major works in philosophy, psychology, and economics, from Maslow (1946) to Stiglitz, Sen, and Fitoussi (2009) and beyond. The list includes aspects that have

Framework.” As part of its Better Life Initiative, the OECD has held several conferences on “Measuring Well-Being for Development and Policy Making” since May, 2011; the organization’s Better Life Index website (www.oecdbetterlifeindex.org) invites visitors to rate 36 countries by interactively creating indices from 11 dimensions of well-being that include SWB. Moreover, this interest is not limited to rich, Western countries; indeed, Bhutan is considered the pioneer of Gross National Happiness, a concept conceived there in the 1970s.

⁴For prominent—and conflicting—proposals to the UK Office of National Statistics, see Dolan, Layard, and Metcalfe (2011) and Deaton et al. (2011).

been considered as fundamental, as well as broader, “combination aspects” that may single-handedly capture many fundamental aspects. While it includes SWB measures widely used by economists (e.g., happiness and life satisfaction), it also includes other items, such as goals and achievements, freedoms, engagement, morality, self-expression, relationships, and the well-being of others. While far from exhaustive, our list represents, as far as we know, the most comprehensive effort to date to construct such a compilation (cf. Alkire 2002).

Next, as described in Section III, we design and conduct a survey to estimate a vector proportional to the vector of marginal utilities $D_w u(\mathbf{w})$. We present more than 4,600 Internet survey respondents—a demographically diverse (albeit not representative) sample of the US adult population—with sets of hypothetical-choice scenarios. We first provide detailed scenario instructions and an example. Then, in each scenario we elicit respondents’ stated preference between two options that differ only on how they rate on a small set of aspects. For example:

Imagine you are **making a personal decision**, and that you face a choice between two options: Option 1 and Option 2. The two options are predicted to have different effects over the next four years but to have the same effects after that. The table below lists these predicted differences in the next four years. Please assume that anything not listed in the table would be marked “about equal” if it were listed.

[Click here to see the instructions again](#)

	OPTION 1					OPTION 2		
	much higher	somewhat higher	slightly higher	about equal		slightly higher	somewhat higher	much higher
how happy you feel							X	
you not feeling anxious			X					

Between these two options, which do you think you would choose?

OPTION 1			OPTION 2		
Much prefer Option 1	Somewhat prefer Option1	Slightly prefer Option 1	Slightly prefer Option 2	Somewhat prefer Option 2	Much prefer Option 2
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

In our estimation procedures, the dependent variable is the response to the choice question above, and the independent variables are the relative ratings of the aspects (“X’s”) in the table above. Because we randomly assign which aspects vary between the options and by how much, we can identify the relative marginal utility of each aspect. We call this stated-preference survey for estimating relative marginal utilities an *SP survey* to distinguish it from an *SWB survey* that would measure individuals’ *levels* of aspects of well-being.⁵ We highlight differences between our specific

⁵Broadly speaking, our SP survey can be viewed as an application of conjoint analysis (Green and Rao 1971). In the context of assessing welfare, our survey design is closely related to the method proposed in Adler and Dolan (2008), who argue that policymakers’ weighting of different aspects of well-being should be informed by how survey respondents rank alternative “lives” that vary in the aspects. They illustrate their method with an exploratory study of 72 undergraduate respondents and four aspects: income, health, happiness, and life expectancy. Relatedly,

SP-survey implementation and the theoretical ideal that we anticipate governments could approximate more closely.

Section IV presents our main survey findings. Using personal-choice scenarios similar to the one above and pooling across all of our respondents we find, among other things, that while commonly measured aspects of well-being such as happiness, life satisfaction, and health are indeed among those with the largest relative marginal utilities, other aspects that are measured less commonly have relative marginal utilities that are at least as large. These include aspects related to family (well-being, happiness, and relationship quality), security (financial, physical, and with regard to life and the future in general), values (morality and meaning), and having options (freedom of choice, and resources). Using policy-choice scenarios, in which respondents vote between two policies that differ in how they affect aspects of well-being for everyone in the nation, we continue to find the patterns above and in addition find high relative marginal utilities for aspects related to political rights, morality of others, and compassion toward others, in particular the poor and others who struggle. While we find some differences across demographic-group and political-orientation subpopulations of our respondents, most of these main results hold across the subpopulations we examine.

We present a long list of robustness checks in Section V. These include exploring the sensitivity of our findings to: estimating alternative econometric specifications; excluding alternative candidates for (non-fundamental) combination aspects from the estimation; examining subsets of our respondents based on the time they took to complete the survey; and varying survey design elements that we randomly manipulated.

In Section VI we return to the two practical questions above—which items to include on an SWB survey, and how to weight them to construct an index—and discuss potential solutions, both in theory and in practice. We outline directions for thinking about implementation challenges such as avoiding double-counting and reducing the number of survey questions.

Our paper contributes to a long line of research on social welfare measures, recently surveyed by Fleurbaey (2009) and Fleurbaey and Blanchet (2013).⁶ Our approach has the appealing feature that it accommodates several traditions that are often considered conflicting. It is, at the same time, “super-liberal” (Fleurbaey 2009) since we weight aspects based *solely* on our respondents’ stated preferences; “welfarist” (Sen 1979a) since utility can be viewed as our exclusive criterion for judging well-being;⁷ and an application of Sen’s (1985, 1992) “capabilities” approach since the vector \mathbf{w} can be considered to include both functionings (achieved states, e.g., material standard of living, health) and capabilities (opportunities for such

Adler (2013) proposes a conceptual framework similar to ours, viewing \mathbf{w} as a hybrid of “mental” (e.g., emotions) and “nonmental” (e.g., freedoms) aspects. He encourages using stated-preference surveys for learning about $u(\mathbf{w})$.

⁶Fleurbaey’s (2009) title, “Beyond GDP: The Quest for a Measure of Social Welfare,” inspired the title of our paper.

⁷Sen (1979a, p.468) defines welfarism as the imposition of the following “informational constraint”: “[t]he judgment of the relative goodness of alternative states of affairs must be based exclusively on, and taken as an increasing function of, the respective collections of individual utilities in these states.” Sen criticizes this notion and argues that information beyond individual utilities may be needed for making moral judgments. To partially address this concern, our approach attempts to include important specific examples of such information as additional arguments in the utility function (i.e., as elements of \mathbf{w}).

achievements, e.g., freedoms).⁸ Indeed, as described above, we find that our respondents put a large weight on capabilities, especially in policy scenarios.⁹

The central assumption underlying our SP-survey methodology is that a person's stated preference in our abstract scenarios is an unbiased measure of her ("true") preference.¹⁰ This assumption is surely wrong; indeed, there are known ways in which stated preference is biased relative to incentivized choice, for example when one choice option is viewed as more socially desirable (Camerer and Hogarth 1999; Ding, Grewal, and Liechty 2005). Nonetheless, we believe it is more attractive to rely on what people's own stated preferences suggest about what they themselves care about than to paternalistically rely on the opinions and introspections of "experts" (such as researchers and policymakers) regarding which aspects to track and how to weight them. Moreover, some of the objections to using stated preferences as if they are descriptive of incentivized choice may have less force when stated preferences are used normatively. In particular, while hypothetical choices in abstract scenarios may elicit meta-preferences (the preferences people *want* themselves to have), rationalized preferences (more deliberated, internally consistent preferences), or otherwise laundered preferences (e.g., omitting "dirty" preferences such as racism), it is sometimes argued that these are more relevant for evaluating welfare than the preferences that describe actual behavior. For example, our SP survey may put respondents in a deliberative frame of mind, causing them to weight emotional factors less than they do in "real life"—but doing exactly that is common prescriptive advice for avoiding emotion-induced mistakes (e.g., Camerer et al. 2003, pp. 1238–1240). We discuss related points when interpreting our results (in Section IV).

In Section VII, we mention some concerns regarding tracking well-being with surveys that we do not address. We conclude by discussing possible extensions of our approach, and we point to a few readily actionable steps suggested by our findings. Throughout, we highlight the limitations of our specific implementation and point out promising directions for further developing the agenda we propose.

I. Theoretical Framework

We start with the standard framework for aggregating an individual's consumption of different commodities. This framework underlies empirical expenditure- and income-based measures of well-being, including GDP; below we adapt it for conceiving a well-being index. We then discuss how this index-oriented framework

⁸From the capabilities-approach perspective, our empirical effort is an attempt to address the "index problem," namely the problem of choosing weights for functionings and capabilities. This problem has been considered a key obstacle to systematically applying the capabilities approach (Fleurbaey 2009) and is in general central to the construction of multidimensional well-being indices (Decancq and Lugo 2013). Although Sen initially opposed combining measures of capabilities and functionings into an index, he later seems to have become more sympathetic (Micklewright 2001).

⁹This also gives some support to Rawls' (1971) contention that most people would prioritize basic rights ahead of other goods.

¹⁰While we interpret our results in light of this assumption, we believe that our empirical contribution can be useful also in the context of alternative theoretical perspectives, such as that of Fleurbaey, Schokkaert, and Decancq (2009), who single out a specific survey question—a life satisfaction question—and propose to use it as the primary source of information about preferences, and those explored by Decancq, Van Ootegem, and Verhofstadt (2011), who preselect nine survey questions and evaluate alternative weighting schemes by directly asking survey respondents how important they consider each of the nine to be. In particular, our approach can inform researchers regarding the link between different survey questions and stated preferences.

can be viewed more generally as an application of choice theory, with preferences elicited via our SP survey. We use this more general perspective to highlight the assumptions underlying the construction and use of our proposed index.

In the consumption context, an agent's well-being is represented by a continuously differentiable utility function $u(\mathbf{c})$. The vector, $\mathbf{c} = (c_1, \dots, c_M)'$, represents the quantities of M market goods. Following a change in the consumption vector, $\Delta \mathbf{c}$, the change in utility can be approximated up to an arbitrary multiplicative scale as

$$(1) \quad \Delta u \approx (D_{\mathbf{c}}u(\mathbf{c}))' \cdot \Delta \mathbf{c} = \sum_{m=1}^M \frac{\partial u(\mathbf{c})}{\partial c_m} \Delta c_m \propto \sum_{m=1}^M p_m \Delta c_m.$$

The proportionality follows because, at the optimum, as long as the consumer chooses a strictly positive amount of each good c_m , each marginal utility $\frac{\partial u(\mathbf{c})}{\partial c_m}$ is equal to a Lagrange multiplier times the market price p_m . By fixing prices, $\bar{p}_1, \dots, \bar{p}_M$, at their levels in some base period and measuring the agent's consumption vector over time, the government can track a quantity index of real consumption, $\sum_{m=1}^M \bar{p}_m c_m$. For small changes in \mathbf{c} , changes in this index are approximately proportional to changes in utility.

Perhaps the biggest limitation of this consumption-based approach is that it only considers a narrow set of determinants of well-being. To broaden its scope, we follow other researchers in shifting attention away from standard consumption goods (for example, rice, TVs, train rides) and toward more fundamental aspects of well-being (for example, health, emotional states, freedoms). This approach is intended to be more general in that these fundamental aspects include all objects of ultimate desire for individuals regardless of which specific consumption goods are in an agent's choice set at a given time and place. In this framework, consumption matters for well-being through its effects on these more fundamental aspects of well-being, but nonconsumption determinants of well-being are also accounted for via their effects on these fundamental aspects.¹¹

Consider the utility function $u(\mathbf{w})$, where $\mathbf{w} = (w_1, \dots, w_J)'$ represents the quantities of J fundamental aspects. Analogously to the consumption formula above (equation (1)), a change in utility resulting from a change in the fundamental aspects can be approximated as

$$(2) \quad \Delta u \approx (D_{\mathbf{w}}u(\mathbf{w}))' \cdot \Delta \mathbf{w} = \sum_{j=1}^J \frac{\partial u(\mathbf{w})}{\partial w_j} \Delta w_j.$$

Instead of measuring the agent's consumption vector \mathbf{c} , the government would measure her fundamental-aspects vector \mathbf{w} ; and instead of tracking a quantity index for standard consumption goods, the government would track $\sum_{j=1}^J \frac{\partial u(\mathbf{w})}{\partial w_j} w_j$, with the marginal utilities fixed at a base period. We term $\sum_{j=1}^J \frac{\partial u(\mathbf{w})}{\partial w_j} w_j$ the agent's *well-being index*. Since the marginal utilities are defined only up to an arbitrary

¹¹ This perspective is essentially a variant of consumer theory in which preferences are defined over goods' characteristics (Lancaster 1966); the fundamental aspects can be thought of as the ultimate underlying characteristics that people care about when making choices.

multiplicative constant, so is the index. From the perspective of this theoretical framework, the purpose of an SWB survey is to measure the w_j s.

Because there are no observable prices that can be used in place of the marginal utilities—by which here and henceforth we mean the *relative* marginal utilities—the government will need direct marginal-utility estimates in order to calculate the agent's well-being index. The purpose of our SP survey is to demonstrate a method for generating such estimates. We envision governments applying such a method on nationally representative samples and, because the marginal utilities may change as \mathbf{w} changes, doing so on a regular basis (just as prices are currently remeasured on a regular basis).

This description of our theoretical framework highlights the analogy between our proposed index and standard consumption-based indices. Our framework, however, is more general. Indeed, our SP survey could in principle be extended to elicit an individual's entire indifference map by designing the survey to elicit stated preference between *every* pair of (explicitly and fully listed) vectors of conceivable responses to an SWB survey. Regardless of the shape of preferences, ordinal welfare comparisons could then be made between any of the individual's SWB-survey occasions. In practice, however, such an unrestricted approach is unimplementable with more than a handful of aspects for two reasons. First, even if assuming transitivity and monotonicity eliminated the need to directly compare *every* pair of aspects, the number of pairwise comparisons required on the survey would grow exponentially with the number of elements in \mathbf{w} . Second, each such pairwise comparison would require respondents to compare two long vectors of aspects, arguably an excessive cognitive demand on survey respondents. Moreover, even if the list of aspects were short, stated-preference data may be less reliable the farther the vectors to be compared are from the respondent's current \mathbf{w} because it may be more difficult for respondents to introspect regarding distant, unfamiliar situations.

Our approach in this paper—and the resulting design of our SP survey detailed in Section III below—can be understood as a more practical version of this generalized procedure, whereby we restrict ourselves in two ways. First, we elicit preferences only locally at the status quo, where we think stated-preference data are most reliable.¹² Second, we assume that preferences can be represented by a continuously differentiable utility function, and hence are locally linear. Local linearity makes the SP survey more feasible: the absence of complementarities allows us to alleviate respondents' cognitive task by presenting them with a series of pairwise comparisons that involve a trade-off between as few as two aspects at a time deviating from current \mathbf{w} ; and the fact that the functional form of local preferences is known reduces the number of SP-survey questions needed to estimate the (local) indifference surface. The continuous-differentiability assumption underlies the discussion surrounding equation (2) above: as long as \mathbf{w} remains within a small neighborhood around the \mathbf{w} prevailing when an SP survey was conducted, the well-being index can be used to make welfare comparisons between SWB-survey occasions.

¹² While our scenarios do not explicitly frame the choice as being relative to respondents' current aspect levels, we believe that this is the natural interpretation (see language in the example scenario in the introduction). One could more explicitly instruct respondents to interpret the choice this way by simply adding a few words to the scenarios' preamble.

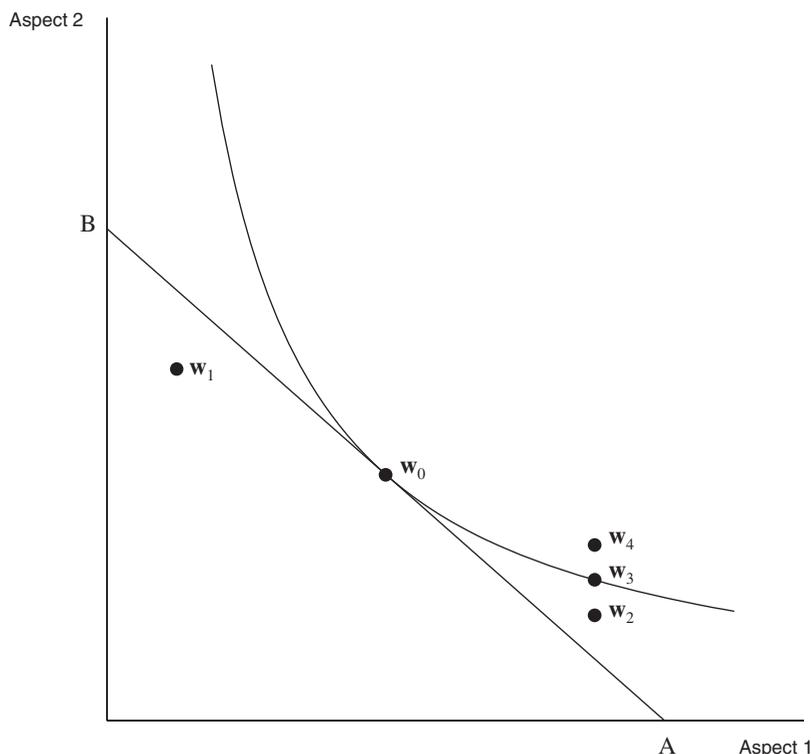


FIGURE 1. PARTIAL WELFARE ORDERING FROM NONLOCAL CHANGES IN THE WELL-BEING INDEX

Such a local welfare ordering may often be sufficient, as many individuals may have reasonably stable w s for long periods of time. However, under some circumstances w may drift away (as time goes by) or even jump discontinuously (perhaps due to a change in policy or in personal circumstances). Between a w within the neighborhood where an SP survey was conducted and a w outside that neighborhood, the index does not provide a complete welfare ordering. Nonetheless, under additional assumptions on preferences, the index could provide a *partial* welfare ordering. For example, suppose that preferences are convex. Figure 1 is adapted from Sen (1979b) who discusses how convexity generates a partial ordering in the context of a real consumption index. The line AB—the tangent to the indifference curve going through w_0 , whose slope is in our case estimated with an SP survey at w_0 —partitions all bundles into a southwest set and a northeast set. Using the weights calculated from an SP survey at w_0 , the index values decrease from w_0 to every aspect bundle in the southwest and increase from w_0 to every bundle in the northeast. Thus if the index value decreases from w_0 to any bundle (e.g., w_1 in the figure), no matter how far away, then due to convexity, we can conclude that the individual is worse off. If the index value *increases* from w_0 , however, the individual could be worse off (w_2 in the figure), indifferent (w_3), or better off (w_4). If a new SP survey were conducted at the new bundle, then it is possible that w_0 would lie in the new southwest set, completing the ordering by revealing that the new bundle is unambiguously better than w_0 . It may also be possible to make a partial ordering more complete by applying transitivity to a sequence of unambiguous pairwise comparisons.

In summary, with a given set of weights estimated at \mathbf{w}_0 , our index provides a complete local ordering, and a partial global ordering, of \mathbf{w} 's. Locally, increases and decreases in the index are both interpretable as reflecting welfare changes, while globally, comparisons require more involved logic such as that above.

Although in this paper we focus primarily on survey-based measures of the fundamental aspects that comprise \mathbf{w} , our theoretical framework also applies when some or all of the aspects are measured objectively. For example, some dimensions of health could be measured with physiological tests, and, arguably, some freedoms could be quantified. Indeed, widely used indices such as the UN's Human Development Index and Okun's "misery index" (the sum of the inflation and the unemployment rates) consist entirely of objective measures that are combined by using ad hoc weights. Regardless of whether the components of \mathbf{w} are measured objectively or subjectively, we propose using a data-driven method for estimating the marginal utilities; our SP survey is an attempt at implementing such a method. Even with objective measures that have observable prices, directly estimating the marginal utilities may be considered an alternative to making the assumptions necessary for using prices as weights.

While our theory above focuses on constructing a well-being index for a single agent, in order to construct a national index in practice, governments would need to, first, construct indices for many people and, second, aggregate them. To construct indices for many people, a method such as our SP survey could be used to estimate marginal utilities for each person in a representative sample. Since doing so may require a long (and hence potentially expensive) survey—especially if \mathbf{w} includes many aspects—in practice pooling data within respondent subpopulations may sometimes be a required compromise. Given our own very limited resources, when demonstrating our methodology in Section IV our main specification pools data across all respondents—something that governments should avoid, as we discuss in Section VI.

How to aggregate utility across individuals is a central question of welfare economics and an active area of research in the literature on social choice. Although this paper focuses on constructing an individual-level well-being index, in principle our framework and empirical contribution are compatible with a variety of approaches to aggregation, as we briefly discuss in Section VII.

II. Aspects of Well-Being

A major obstacle to any real-world application of our theoretical framework is that no one knows which fundamental aspects comprise the vector \mathbf{w} . Indeed, different authors have proposed different sets of aspects as important components of well-being. Our treatment of aspects as arguments of utility requires that for our purposes, any proposed set must be *exhaustive*, i.e., include *all* arguments of preferences; and it must comprise aspects that are *nonoverlapping*, i.e., that are conceptually distinct.

Our approach in this paper is to construct as comprehensive a list of candidate fundamental aspects as we practically can. We additionally include what we believed would be broader, non-fundamental "combination aspects"—by which we mean facets of well-being that are not themselves fundamental but contain information regarding multiple fundamental aspects—which might capture more of the

variation in u . Our comprehensiveness has three advantages. First, it reduces the risk of missing important components of \mathbf{w} , thereby making the list closer to exhaustive. Second, it allows us to minimize the influence of our own *ex ante* beliefs on the set that emerges as important from our analysis.¹³ Third, it renders our results as broadly useful as possible, since different researchers can focus on the subset of our results that pertains to the aspects they believe comprise \mathbf{w} —or to those they happen to have data on in an existing social survey. On the other hand, our attempt to be comprehensive has the drawback of increasing the likelihood that different aspects on our list overlap with each other. Our analysis in Section V, in which we exclude some potentially overlapping aspects from the analysis, suggests that such possible overlap does not meaningfully affect our marginal utility estimates. However, in selecting which aspects to include in a well-being index, a method will be required for ensuring that fundamental aspects are not double-counted. We summarize a proposal for such a method in Section VI; we discuss it in more detail and formalize it in the online Appendix.

In this section we summarize our method for compiling our master list of 136 aspects of well-being (see Table 2 for a version of the list). The list, as well as much more detail regarding the process of compiling it, dividing it into different sub-lists, and creating different versions of it, is available in the online Appendix.

Our list draws from six classes of survey measures. First, we include single-question SWB measures, modeled after the SWB questions most commonly asked in large-scale social surveys (for example, those asked in or proposed for the UK survey discussed above). These include mostly measures that are considered *evaluative/cognitive* (e.g., life satisfaction)—i.e., measures based on questions that may trigger respondents to engage deliberative cognitive processes in order to evaluate their life—and those that are considered *hedonic/affective* (e.g., felt happiness)—i.e., measures that are aimed at eliciting respondents' positive or negative feelings.

Second, we include measures based on items in multi-question survey measures of SWB. These are primarily drawn from scales commonly used in psychology, such as the PANAS (Positive and Negative Affect Scale).

The third and largest class contains aspects of well-being proposed by prominent economists, psychologists, and philosophers that are not typically elicited on large surveys. We drew on proposals from the Stiglitz Commission (Stiglitz, Sen, and Fitoussi 2009), the systematic compilation effort by Alkire (2002), as well as many classic sources (e.g., Maslow 1946; Sen 1985; Nussbaum 2000), and more recent contributions (Diener and Seligman 2004; Loewenstein and Ubel 2008; Graham 2011). A complete list of the works we reviewed, including references by aspect, is given in the online Appendix. The many aspects in this class include some that can be understood as capabilities, i.e., access to resources, choice sets, and freedoms. The aspects also include some that would be considered *eudaimonic* SWB measures (for example, having a meaningful life; Ryff 1989). Such measures relate to human thriving or flourishing; the Greek word *eudaimōn* is commonly translated to mean

¹³That said, in compiling the list we draw exclusively on English-speaking sources, introducing a cultural bias—but one compatible with our respondents'. In addition, "dirty" aspects (such as racist objectives) are typically absent from our sources, but as we discuss in Section IV, the few aspects on our list that could be considered "dirty" (e.g., social status) are ranked very low by our SP-survey respondents anyway.

happy in an Aristotelian sense, i.e., as associated with having a meaningful, valuable, or worthwhile life.

The fourth class of measures resulted from our own introspection and discussion. Some of our proposed additions were confirmed to be important in our past work (Benjamin et al. 2012), and others resulted from extensive discussions among ourselves and with colleagues. Some reflect our attempt to break down important parts of life into more fundamental aspects. For example, while many writers have proposed that religion is important for well-being, we refined “religion” into 15 aspects of well-being that may help explain the value of religion but that are also valued by many nonreligious people (e.g., “you having people around you who share your values, beliefs, and interests”).

While these four classes of measures represent our attempt to include fundamental aspects in our list, our fifth and sixth classes serve different purposes. The fifth class represents our attempt to formulate combination-aspect survey questions that, if broad enough, might even serve as good empirical proxies for u itself, thus possibly obviating the need for an index. These include novel (in wording, not necessarily in concept), evaluative well-being measures, such as “how much you like your life” and “the overall well-being of you and your family.”

Finally, as a sixth class, we crafted survey versions of “objective” indicators that are widely used as measures of well-being, such as the rates of GDP growth, inflation, and unemployment, or income inequality. The weights respondents put on such measures can serve as a benchmark against which we can compare measures from our first five classes.¹⁴

After compiling an initial list, we revised it according to several criteria. To make its length more manageable, we combined similar items into single measures (but we preserved commonly-used survey questions close to their original form). To reduce subject confusion and response error, we oriented all items so that rating higher would conventionally be considered desirable; for example, “not feeling anxious.” We further edited items in order to use vocabulary that would be understandable by most respondents in a national sample.

The final list of 136 aspects includes 113 “private good” aspects—relating to an individual’s own well-being (e.g., “your health”)—and 23 “public good” aspects (also labeled *public*-aspects)—relating to an entire society’s well-being (e.g., “equality of opportunity in your nation”). Among the private-good aspects, we distinguish between what we label *you*-aspects—108 that pertain to the respondent but could in principle pertain to everyone (e.g., “your health”)—and *you-only*-aspects—five that pertain to the respondent and could *not* meaningfully and distinctly pertain to everyone at the same time, for example due to their being inherently relative (e.g., “your social status”).

¹⁴In principle, we could have additionally included receipt of different amounts of money as an aspect on our list. Doing so would have allowed us to scale the estimated marginal utilities of the nonmoney aspects, converting them to dollar units, in turn enabling one to sum equivalent variation across individuals. In practice, however, this approach would have significantly complicated our current (purposefully simplified, proof-of-concept) survey design. For example, dollar amounts would require their own quantitative scales that could not be made easily comparable to the rating scales of all other aspects without considerable design changes (see the discussion in Section IIIC below).

Finally, from each of the 108 *you*-aspects we constructed two additional aspect versions: an *everyone*-aspect that pertains to everyone in a nation (e.g., by replacing “your health” with “people’s health”); and an *others*-aspect that pertains to typical others (e.g., “others’ health”). In the next section, we explain the purpose of those aspect versions and discuss them further.

III. Survey Design

The core of our online SP survey consists of 30 hypothetical-choice scenarios, one per screen. They are preceded by a screen of detailed instructions and followed by a multiscreen exit questionnaire.

A. Scenario Screens

An example scenario, as it appears on the screen, was reproduced in the introduction. Each such screen has three components. First, the *preamble* frames the scenario as a choice between two options, neutrally labeled “Option 1” and “Option 2,” that have different consequences over the next four years. Second, the *aspect table* describes the difference in consequences between the two options. Finally, the *choice question* elicits a participant’s stated preference between the two options.

Preamble.—The preamble appears in one of two versions. The first version (reproduced in the introduction) introduces *personal-choice* scenarios. Since much of the discourse regarding SWB surveys is focused on private-good aspects, personal choice seems the relevant setting for eliciting these aspects’ weights through respondents’ pairwise decisions that trade them off.

The second preamble version introduces *policy-vote* scenarios: the opening clause “Imagine you are **making a personal decision**” is replaced with the clause “Imagine that you and everyone else in your nation are **voting on a national policy issue**.”¹⁵ Policy-vote scenarios have two purposes. First, as with standard public goods, our 23 *public*-aspects cannot typically be affected by one individual’s personal choice—but are routinely traded off in policy-vote contexts. Second, even for *you*-aspects, if a national SWB survey is to be used for evaluating policy, it may be useful to elicit the relative weights also in a setup where the aspects pertain to everyone (and are traded off by policy in the same way for everyone). Due to other-regarding preferences, for example, these *everyone*-weights could differ from their corresponding *you*-weights elicited in personal-choice scenarios. While our empirical effort is focused on personal scenarios, we also explore such personal-versus-policy comparisons below.

The rest of the preamble is identical across all scenarios. Designed to elicit participants’ “single-period” utility, it explicitly limits the duration of the predicted

¹⁵We chose this voting frame for two reasons. First, it is designed to imitate a situation that, while hypothetical, is as close as possible to the policy choice setup people have real-world experience with; while most people are never in a position to actually choose a policy that would get implemented, people often face choices over real policies in the voting booth. Second, the frame is designed to elicit participants’ *own* preferences. While these preferences might well incorporate a concern for others, we did *not* want the choices to put extra weight on others out of concern that these others would be left out of the decision-making process. For this reason the question explicitly states that “you and everyone else in your nation” are voting.

difference between the options. While four years is a somewhat arbitrary duration, it does not seem unreasonable as a time frame for assessing policy (for example, it is the length of the term of the US President) as well as personal choices. The preamble ends with a sentence that effectively asks participants to imagine that anything not explicitly stated to differ is held constant.

Aspect Table.—Each row of the aspect table compares the two options in terms of one aspect, with an “X” positioned to indicate that either Option 1 or 2 rates “much higher,” “somewhat higher,” or “slightly higher” on that aspect. The “about equal” column in the middle never contains an “X”; it serves as a reminder to participants that unlisted aspects are to be considered as affected equally by the two options.

Personal scenarios draw aspects randomly from the set of 113 private aspects (which consists of 108 *you-* and 5 *you-only-*aspects). The policy scenarios we analyze in this paper instead draw aspects randomly from the set of 108 *everyone-* and 23 *public-*aspects; these 131 aspects are effectively public goods because they affect everyone in the same way. Each respondent faces, in random order, 11 personal scenarios, 5 of these policy scenarios, and 14 additional exploratory versions of policy scenarios that we do not analyze in this paper.¹⁶

The number of aspects (or rows) in a scenario’s aspect table is randomly drawn from the set {2, 3, 4, 6}. While a shorter table may be easier for respondents to read and think through, a longer one improves statistical power for identifying marginal utilities.

The rating of each aspect—i.e., the location of the “X” in each row—is randomly drawn from the six feasible ratings. However, we place two restrictions on the combination of ratings within a scenario. First, scenarios with an even number of aspects must be *balanced*: exactly half of the aspects favor Option 1 and exactly half favor Option 2. Second, scenarios with four or six aspects must additionally be *symmetric*: each rating in favor of Option 1—i.e., “much higher,” “somewhat higher,” or “slightly higher”—is matched by a rating of the same intensity in favor of Option 2 on another aspect.¹⁷

Choice Question.—The choice response scale is identical across all scenarios. It is designed to elicit intensity of preference on a six-point scale (“Much prefer Option 1,” “Somewhat prefer Option 1,” “Slightly prefer Option 1,” “Slightly prefer Option 2,” etc.). To discourage “lazy” responses, we omitted an “indifferent” option; our “slightly” options are intended to allow for nearly-indifferent choices.

¹⁶These additional versions are designed to explore issues related to other-regarding preferences that are beyond the scope of this paper. (For example, to what extent do people’s votes on policy reflect their willingness to sacrifice their own utility for increasing others’ utility?) These policy scenario versions draw aspects from sets that consist of different combinations of *you-*, *you-only-*, *others-*, and *public-*aspects.

¹⁷We require two-aspect scenarios to be balanced because, otherwise, half of them are expected to elicit a trivial choice, as one of the options would rank higher on all aspects. By placing no restrictions on three-aspect scenarios, we allow such trivial cases to occur (with 25 percent probability), and we use them as a secondary robustness check to identify respondents who might have answered at random (see footnote 29 below). Finally, we require four- and six-aspect scenarios to be balanced and symmetric because our pre-tests suggested that otherwise, participants faced with these longer scenarios may adopt “visual shortcut” heuristics—such as choosing an option that ranks better on most aspects—and pay less attention to the identities of the aspects. We compare results estimated separately from two-, three-, four-, and six-aspect scenarios in Section V and in the online Appendix.

B. Instructions and Questionnaire

The instructions screen is reproduced in the online Appendix. Respondents could reopen it from every scenario screen by clicking a hyperlink. It includes an example aspect table that is more complex than that shown in the introduction, illustrating and explaining more possibilities. The instructions emphasize a number of scenario design points, including: the distinctions between personal-choice and policy-vote scenarios; that the two options differ *only* on the consequences listed in the aspect table; and that “The items and their rankings in the tables are randomly chosen by the computer so that we [the researchers] learn as much as possible from your choices.”¹⁸

In addition, the instructions explain that in each row of the aspect tables, one word is emphasized in boldface type (in the example in the introduction, **you** and **your**). Participants are asked to “pay careful attention to the emphasized words, and interpret a consequence with the following emphasized words as affecting the following people.”¹⁹

you/your: affects only you (and, when stated, your family or close friends).

others/others’: does not affect you, your family, or your friends—but does affect other people. The table indicates the average effect on other people in your nation.

nation/society/people: affects everyone in your nation (including you, your family, and your friends).

world/humanity: affects everyone in the world.

Once participants complete the scenarios, they fill out an exit questionnaire. We ask participants, in both a multiple-choice question and an open-ended question, whether they understood what we were asking them to do (in the multiple choice, 92.5 percent answered “always” or “mostly”; 7.5 percent answered “not really”). We also ask basic demographic questions, as well as questions about ideology, political party affiliation, and religiosity.

C. Design versus Theory

As discussed above, our SP survey is intended as a first-pass demonstration of feasibility. As such, it is a considerably simplified version of the theory-based ideal that should guide governments as they design their surveys. We now briefly describe what a theoretically ideal survey might involve, and we highlight a few key deviations of our SP survey from that ideal.

In an ideal application of our theory, an SP-survey component would follow an SWB-survey component in a single, combined survey. The SWB component

¹⁸This randomization disclosure is intended to prevent respondents from perceiving random combinations of aspects as reflecting an intentional design decision to construct a particular scenario. Such mistaken inferences could give rise to undesired experimenter-demand effects.

¹⁹We included the bolding and this “legend” (reproduced here as it appears on the screen) because pre-tests without the bolding indicated that participants might quickly skim the aspect tables, mistakenly assuming that, for example, all aspects apply to only them personally. The bolding scheme allows us to keep the aspects short and simple and at the same time visually clear and coherent. The legend provides respondents with a quick reference that clarifies the bolded words’ intended meaning.

would elicit a respondent's current aspect levels (i.e., her \mathbf{w}). Each of the J fundamental-aspect questions would have its own scale: while a subjective scale with verbal labels may be a natural choice for some aspects, for other aspects it may be possible to use an objective scale with quantitative units. After respondents report their current \mathbf{w} in the SWB-survey component, the SP component would elicit their choices between pairs of options, where each option involves a small change from the current \mathbf{w} ; these small changes would be spelled out using each aspect's own units. With sufficiently many such choices, each respondent's vector of marginal utilities could then be estimated.

Our SP survey deviates from this ideal in a number of ways. First, it is a stand-alone survey, rather than the second component in a combined SWB+SP survey. We concentrate our efforts on a stand-alone SP survey since SWB surveys are already conducted routinely and do not require a demonstration of feasibility. Second, our survey uses verbal, relative rating scales that are identical for all aspects ("much," "somewhat," and "slightly higher") rather than stating differences using each aspect's own units. We do this to simplify and streamline the survey and its instructions. It also allows us to interpret relative marginal-utility estimates as informative regarding aspects' relative "importance" because the aspect ratings are put in units that are arguably comparable, namely respondents' judgment of what constitutes "slightly," "somewhat," or "much" more of the aspect (but see Section VIB for a discussion of how, with additional data, aspects could be compared in terms of predictive power for the index). Third, our survey elicits few SP questions from each respondent, and in our marginal utility estimates below we pool the data across respondents. Doing so keeps the survey short—a crucial feature given our convenience sample—but theoretically would only be justified if respondents not only use the scales the same way but also have the same local slopes of indifference surfaces. We return to this last point in Section VIC.

IV. Empirical Results: Marginal Utility Estimates

Our SP-survey respondents were recruited during December 2011 by Clear Voice Research, a private firm that invites individuals to "start sharing their voice" and "make a little money" by participating in online surveys. We aimed at a sample that, although not a random sample, would resemble the adult (20+) US population on the demographics listed in Table 1.

A. Respondent Demographics

Table 1 reports the demographic distribution of the 7,391 respondents who began our survey (the "All" column); the 5,397 who completed it ("Completes"); and the 4,608 included in our main analysis below ("Primary sample").²⁰ This primary sample excludes respondents who completed the survey in less than eight minutes;

²⁰Due to a programming error, the first 1,936 primary-sample respondents faced scenarios in which aspects were unintentionally drawn from only 108 of the intended 113 personal and 131 policy aspects. As a result, we have more data—and tighter estimates—on some aspects. Excluding these 1,936 early respondents has very little effect on our main estimates (see online Appendix Table WA1).

TABLE 1—RESPONDENT DEMOGRAPHICS

		All (<i>N</i> = 7,391)	Completes (<i>N</i> = 5,397)	Primary sample (<i>N</i> = 4,608)	Census etc.
Marital status	Married	59.3	58.9	59.8	48.8
	Never married	24.3	25.5	23.4	32.1
	Other	16.4	15.6	16.8	19.1
Highest education level completed	High school grad	23.2	21.9	21.4	42.9
	Some college	40.7	39.8	41.0	28.9
	Bachelor's degree	24.3	25.4	25.5	17.7
	Graduate degree	11.5	12.7	12.0	10.4
Age	20–29	19.7	21.1	18.0	18.9
	30–39	18.3	19.5	18.4	17.8
	40–49	19.3	20.0	20.0	19.3
	50–64	20.8	20.4	22.0	26.1
	65 and older	21.9	19.0	21.5	17.9
Income	Less than \$20,000	17.9	18.1	17.4	19.8
	\$20,000–\$39,999	27.9	27.0	27.9	21.7
	\$40,000–\$49,999	10.9	10.2	10.8	8.9
	\$50,000–\$74,999	19.8	20.0	20.5	17.7
	\$75,000–\$99,999	11.3	11.7	11.6	11.4
	\$100,000 and above	12.1	13.0	11.8	20.4
Region	Midwest	23.0	23.2	24.3	21.7
	Northeast	19.8	20.0	19.3	17.9
	South	34.0	33.4	33.8	37.1
	West	23.2	23.4	22.6	23.3
Race	White	76.6	75.2	78.2	63.7
	Black	9.8	10.1	9.6	12.2
	Hispanic/Latino	7.7	8.2	6.4	15.4
	Asian	3.7	4.2	3.5	4.7
	Other	2.2	2.2	2.2	4.0
Household size	1	18.0	17.7	18.2	26.7
	2	35.2	33.8	35.1	32.8
	3	18.8	19.1	18.6	16.1
	4 and above	27.8	29.2	27.9	24.4
Employment status	Employed	53.7	55.9	53.3	57.4
	Unemployed	8.6	8.7	8.8	6.9
	Not in labor force	37.6	35.4	37.8	35.6

Notes: All numbers are percentages. “All”: respondents who began the survey. “Completes”: respondents who completed all scenarios. “Primary sample”: respondents who took at least eight minutes to complete the survey.

Sources: Authors' survey, 2010 American Community Survey, 2010 Census, 2011 Current Population Survey

our robustness analysis in Section V below suggests that including them does not affect our results qualitatively but increases measurement error. The table shows that the three groups are similar on observables.

The rightmost column reports figures from the 2010 American Community Survey, 2010 Census, and 2011 Current Population Survey (see the online Appendix for details). Relative to the US population, our respondents are more likely to be married, college-educated, and white, and less likely to have very high income, be Hispanic, and live alone. Our respondents may of course also differ on unobservables.

B. Personal Choices: Benchmark Specification and Results

Our main results are reported in the “Personal” panel of Table 2. Recall from Section III that each respondent faced 11 personal scenarios where the two choice

TABLE 2—RELATIVE MARGINAL UTILITY ESTIMATES

Aspect	Personal			Policy		
	Coefficient	SE	Rank	Coefficient	SE	Rank
Freedom from corruption, injustice, and abuse of power in your nation				0.39	0.026	0a
The overall well-being of you and your family	0.46	0.016	1	0.33	0.033	3
The happiness of your family	0.43	0.017	2	0.24	0.024	21
Your health	0.42	0.017	3	0.29	0.025	6
You being a good, moral person and living according to your personal values	0.40	0.017	4	0.35	0.025	2
The quality of your family relationships	0.37	0.017	5	0.25	0.024	13
Society helping the poor and others who struggle				0.30	0.024	5a
The morality, ethics, and goodness of other people in your nation				0.29	0.024	5b
Your financial security	0.34	0.017	6	0.28	0.023	8
Freedom of speech and people’s ability to take part in the political process and community life				0.29	0.025	6a
The well-being of the people in your nation				0.29	0.024	6b
Your mental health and emotional stability	0.34	0.016	7	0.25	0.025	15
Your sense of security about life and the future in general	0.33	0.016	8	0.26	0.024	12
You having many options and possibilities in your life and the freedom to choose among them	0.32	0.017	9	0.35	0.034	1
The amount of freedom in society				0.27	0.025	9a
Your sense that your life is meaningful and has value	0.32	0.017	10	0.27	0.023	9
How satisfied you are with your life	0.31	0.017	11	0.18	0.033	53
You feeling that you have enough time and money for the things that are most important to you	0.30	0.017	12	0.21	0.023	32
How much you like your life	0.30	0.017	13	0.19	0.031	46
How peaceful, calm, and harmonious your life is	0.29	0.017	14	0.24	0.024	18
Your nation being a just society				0.25	0.023	14a
Your feeling of independence and self-sufficiency	0.29	0.016	15	0.23	0.024	25
Your pride and respect for yourself	0.29	0.017	16	0.19	0.024	44
Your sense that you are standing up for what you believe in	0.29	0.017	17	0.21	0.025	33
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	0.017	18	0.32	0.025	4
How low the rate of unemployment is in your nation				0.24	0.024	18a
How much you enjoy your life	0.29	0.016	19	0.24	0.025	17
Trust among the people in your nation				0.24	0.023	19a
Equality of opportunity in your nation				0.24	0.023	19b
The extent to which you feel the things you do in your life are worthwhile	0.28	0.016	20	0.24	0.032	23
Your physical safety and security	0.28	0.016	21	0.24	0.023	22
The well-being of the people in the world				0.24	0.024	21a
You “being the person you want to be”	0.28	0.017	22	0.17	0.024	62
Your freedom from being lied to, deceived, or betrayed	0.28	0.017	23	0.30	0.026	5

(Continued)

options differ on 2, 3, 4, or 6 of the 113 personal aspects. Pooling all such scenarios across all respondents, we report results from the following OLS regression:

$$(3) \quad StatedPreference_s = \alpha + \mathbf{AspectRatings}'_s \cdot \beta + \varepsilon_s.$$

Each observation s captures the information from a single scenario faced by a respondent, corresponding to a single survey screen like the example in the introduction.

TABLE 2—RELATIVE MARGINAL UTILITY ESTIMATES (*Continued*)

Aspect	Personal			Policy		
	Coefficient	SE	Rank	Coefficient	SE	Rank
People getting the rewards and punishments they deserve				0.23	0.025	23a
You having people you can turn to in time of need	0.28	0.016	24	0.28	0.024	7
The extent to which you “have a good life”	0.28	0.016	25	0.24	0.032	19
The condition of animals, nature, and the environment in the world				0.22	0.026	25a
You having the people around you think well of you and treat you with dignity and respect	0.27	0.017	26	0.26	0.023	11
How grateful you feel for the things in your life	0.27	0.017	27	0.22	0.032	26
Your sense of control over your life	0.27	0.017	28	0.24	0.023	20
How much love there is in your life	0.27	0.016	29	0.20	0.025	42
How much you appreciate your life	0.27	0.016	30	0.18	0.033	58
How much of the time you feel happy	0.27	0.017	31	0.20	0.025	41
Your sense that things are getting better and better	0.27	0.016	32	0.19	0.032	49
Your sense that you know what to do when you face choices in your life	0.26	0.016	33	0.18	0.023	55
The extent to which humanity does things worthy of pride				0.21	0.024	33a
You having people around you who share your values, beliefs and interests	0.26	0.016	34	0.21	0.024	29
How often you smile or laugh	0.26	0.017	35	0.14	0.024	82
Your ability to dream and pursue your dreams	0.26	0.016	36	0.25	0.024	14
Your chance to live a long life	0.26	0.016	37	0.19	0.025	45
The amount of love in the world				0.20	0.025	37a
How fulfilling your life is	0.26	0.017	38	0.15	0.037	78
How happy you feel	0.26	0.016	39	0.23	0.034	24
How glad you are to have the life you have rather than a different life	0.25	0.016	40	0.15	0.034	73
Your passion and enthusiasm about things in your life	0.25	0.016	41	0.15	0.024	77
You feeling alive and full of energy	0.25	0.016	42	0.15	0.023	76
Your ability to fulfill your potential	0.25	0.016	43	0.21	0.022	28
How low the rate of inflation is in your nation’s economy				0.19	0.024	43a
Your ability to be yourself and express yourself	0.25	0.016	44	0.19	0.023	47
The absence of stress in your life	0.25	0.017	45	0.15	0.024	74
Your ability to keep good perspective in your life	0.25	0.017	46	0.18	0.025	57
Your sense of purpose	0.25	0.017	47	0.21	0.023	35
The amount of order and stability in your life	0.24	0.016	48	0.18	0.023	56
Your freedom from pain	0.24	0.017	49	0.26	0.022	10
You feeling that things are going well for you	0.24	0.016	50	0.21	0.035	34
The quality of your romantic relationships, marriage, love life, or sex life	0.24	0.017	51	0.16	0.026	69
The absence of worry in your life	0.23	0.016	52	0.20	0.024	38
Your sense that you are competent and capable in the activities that matter to you	0.23	0.017	53	0.20	0.024	39
Your physical comfort	0.23	0.015	54	0.12	0.023	88
The amount of order and stability in society				0.18	0.025	54a
How full of beautiful memories your life is	0.23	0.016	55	0.11	0.032	90
Your success at accomplishing your goals	0.23	0.017	56	0.21	0.023	36
Your ability to shape and influence the things around you	0.23	0.016	57	0.19	0.024	50
The rate of economic growth (GDP growth) over time in your nation				0.18	0.026	57a
You feeling that your life has direction	0.22	0.017	58	0.16	0.023	67
How rewarding the activities in your life are	0.22	0.016	59	0.16	0.034	68
You getting the things you want out of life	0.22	0.017	60	0.15	0.033	79
Your sense of optimism about your future	0.22	0.016	61	0.22	0.034	27
The average income of people in your nation (GDP per capita)				0.17	0.024	61a

(Continued)

TABLE 2—RELATIVE MARGINAL UTILITY ESTIMATES (*Continued*)

Aspect	Personal			Policy		
	Coefficient	SE	Rank	Coefficient	SE	Rank
You feeling that you have been fortunate in your life	0.22	0.016	62	0.18	0.033	54
The extent to which your nation does things worthy of pride				0.17	0.024	62a
Your knowledge, skills, and access to information	0.22	0.016	63	0.21	0.023	31
The absence of sadness in your life	0.22	0.017	64	0.20	0.024	37
How often you can feel relaxed instead of feeling your life is hectic	0.22	0.016	65	0.13	0.023	85
Your sense of achievement and excellence	0.21	0.016	66	0.16	0.025	65
Equality of income in your nation				0.16	0.024	66a
The absence of frustration in your life	0.21	0.016	67	0.17	0.025	61
Your freedom from emotional abuse or harassment	0.20	0.017	68	0.25	0.025	16
You not feeling depressed	0.20	0.016	69	0.15	0.025	80
Your ability to have and raise children	0.20	0.017	70	0.18	0.024	59
You feeling that you are part of something bigger than yourself	0.20	0.016	71	0.17	0.024	60
You having many moments in your life when you feel inspired	0.20	0.016	72	0.15	0.025	71
The amount of pleasure in your life	0.20	0.017	73	0.19	0.024	43
Your personal growth	0.19	0.016	74	0.14	0.024	81
The happiness of your friends	0.19	0.024	74a			
How often you are able to challenge your mind in a productive or enjoyable way	0.19	0.016	75	0.15	0.024	75
The absence of anger in your life	0.19	0.017	76	0.19	0.025	51
The quality of your sleep	0.19	0.016	77	0.09	0.025	96
You feeling that you understand the world and the things going on around you	0.19	0.016	78	0.15	0.025	70
Your sense that everything happens for a reason	0.18	0.016	79	0.09	0.024	95
The absence of fear in your life	0.18	0.017	80	0.20	0.026	40
How easy and free of annoyances your life is	0.17	0.016	81	0.07	0.025	98
How desirable your life is	0.17	0.016	82	0.09	0.035	97
Your ability to fully experience the entire range of healthy human emotions	0.17	0.016	83	0.19	0.024	48
Your ability to use your imagination and be creative	0.17	0.017	84	0.16	0.023	66
Your sense of discovery and wonder	0.16	0.016	85	0.11	0.024	91
Freedom of conscience and belief in your nation				0.13	0.024	85a
How close your life is to being ideal	0.16	0.017	86	0.12	0.032	87
Your sense of community, belonging, and connection with other people	0.16	0.017	87	0.21	0.024	30
You not being lonely	0.15	0.016	88	0.17	0.024	64
The total size of your nation's economy (GDP)				0.12	0.024	88a
You feeling that you are understood	0.15	0.017	89	0.19	0.024	52
Your absence of internal conflict (conflict within yourself)	0.15	0.016	90	0.15	0.024	72
The absence of regret you feel about your life	0.14	0.016	91	0.02	0.033	105
You not feeling anxious	0.13	0.017	92	0.09	0.024	94
How interesting, fascinating, and free of boredom your life is	0.13	0.016	93	0.03	0.023	103
You having new things, adventure, and excitement in your life	0.12	0.016	94	0.13	0.024	86
The amount of fun and play in your life	0.12	0.016	95	0.11	0.024	89
Your sense of connection with the universe or the power behind the universe	0.12	0.017	96	0.10	0.025	93
How much beauty you experience in your life	0.11	0.017	97	0.04	0.025	101
Your material standard of living	0.11	0.016	98	0.06	0.026	100
The overall quality of your experience at work	0.10	0.017	99	0.14	0.023	83
You having a role to play in society	0.10	0.016	100	0.17	0.024	63
Your opportunities to participate in ceremonies, cultural events, and celebrations that are meaningful to you	0.09	0.016	101	0.11	0.024	92

(Continued)

TABLE 2—RELATIVE MARGINAL UTILITY ESTIMATES (*Continued*)

Aspect	Personal			Policy		
	Coefficient	SE	Rank	Coefficient	SE	Rank
How often you become deeply engaged in your daily activities (so deeply engaged that you lose track of time)	0.09	0.017	102	0.01	0.023	106
Your rating of your life on a ladder where the lowest rung is “worst possible life for you” and the highest rung is “best possible life for you”	0.09	0.016	103	0.04	0.030	102
The absence of shame and guilt in your life	0.07	0.016	104	0.00	0.025	107
You having a beautiful life story, or a life that is “like a work of art”	0.07	0.016	105	0.13	0.034	84
The absence of humiliation and embarrassment in your life	0.07	0.016	106	0.07	0.025	99
You having others remember you and your accomplishments long after your death	0.04	0.022	106a			
Your ability to “be in the moment”	0.04	0.017	107	-0.02	0.023	108
Your enjoyment of winning, competing, and facing challenges	0.04	0.016	108	0.02	0.024	104
How high your income is compared to the income of other people around you	0.03	0.022	108a			
Your social status	-0.06	0.022	108b			
Your power over other people	-0.09	0.022	108c			

Notes: Personal panel: OLS regression of stated preference on 113 personal aspects and a constant (const. = -0.02, SE = 0.003), using personal choice scenarios ($N = 50,688$). Policy panel: OLS regression of stated preference on the 131 policy aspects and a constant (const. = -0.01, SE = 0.004) using the policy vote scenarios ($N = 23,040$). Standard errors clustered at the respondent level. For the 108 aspects that have versions that appear in both types of scenarios, the text used in the personal choice scenarios is shown; in policy choice scenarios, “people” and “people’s” replace “you and “your.” For each scenario type, the numbers 1–108 are used to rank, by coefficient size, aspects that appear in both scenario types. An aspect that appears in one scenario type receives a rank with a letter: 5a, for example, indicates that the coefficient was between the aspects ranked 5 and 6.

$StatedPreference_s$ encodes the response to the choice question. $AspectRatings_s$, a 113-element vector, encodes the differences between the two options; all of its entries are 0 except for the 2–6 entries representing the aspects on which the options differ. We cluster standard errors at the respondent level.

To the six points on the choice scale (“Much prefer Option 1,” etc.) we assign the six numerical values (-1, -0.47, -0.14, +0.14, +0.47, +1), and to the seven columns in the aspects table (“Option 1 much higher,” etc.) we assign the values (-1, -0.83, -0.75, 0, +0.75, +0.83, +1). As described in Section V below, these numerical values were estimated from the data using a nonlinear ordered probit model, constraining the scales to be symmetric (to economize on parameters) and range from -1 to +1. While we prefer using these estimated scales, misspecifying the scales should have little effect on the estimated aspect coefficients relative to each other. Indeed, as reported in the online Appendix, for the personal scenarios, the correlation between the coefficients reported in Table 2 and those estimated with linear codings (i.e., choice scale: (-3, -2, -1, +1, +2, +3) and aspect ratings: (-3, -2, -1, 0, +1, +2, +3)) is 0.998. We find similarly high correlations (0.99) with coefficients estimated from a probit or logit model where we collapse the choice scale to a binary variable (prefer Option 1 versus Option 2) and use the estimated scales for the aspect columns.

Here we report OLS for maximum transparency, and for simplicity we assume for now that respondents are identical in both their marginal rates of substitution and

their use of the response scale. As reported in Section V below, our results are essentially unaffected when we relax many of the restrictions imposed by this specification. We provide some evidence on heterogeneity across respondent subpopulations later in this section.

We interpret the regression (3) as estimating equation (2), where in our empirical analysis, $\Delta \mathbf{w}$ is the difference in aspect ratings across the two options in a scenario. The β vector estimates a vector proportional to $D_{\mathbf{w}}u(\mathbf{w})$, and ε_s captures response error. We interpret α , which we estimate to be -0.02 (s.e. = 0.003), as picking up a very small respondent bias in favor of “Option 1” despite the fact that the content of the two options is randomly drawn from the same distribution.

The Personal panel in Table 2 reports, for each of the 113 aspect regressors, its coefficient and standard error. The “Rank” column orders the *you*-aspects by coefficient size (1–108) and, additionally, places the 5 *you-only*-aspects relative to these by assigning to them rank numbers with letter suffixes (e.g., “74a” lies between 74 and 75).²¹ Since the independent and dependent variables are coded over ranges of the same length (-1 to $+1$), a coefficient of, for example, $+0.46$ means that on average, changing the relevant aspect from the extreme rating “Option 1 much higher” to the other extreme of “Option 2 much higher” causes choice to move 46 percent of the entire choice scale in the same direction.

Figure 2 summarizes the coefficient and rank information in Table 2 graphically, for the 113 personal aspects (x’s) and the 131 policy aspects (triangles; we discuss these below), sorted by their respective within-panel rank. Of the 113 personal aspects, all but two are positive, almost all statistically significantly so. This confirms that, as intended by our wording of the aspects, an option rating higher on an aspect is, *ceteris paribus*, generally considered preferable. (We discuss the two coefficients that are negative below.) The table and the figure show that the greatest variation in coefficient size across aspects occurs among those at the top (for example, the top ten coefficients range from $+0.46$ to $+0.32$), and at the bottom (the bottom ten range from $+0.09$ to -0.09); coefficients vary more slowly among middle-ranking aspects. The standard errors on the coefficients are typically below 0.017. These features of our estimates should be borne in mind when reading our discussion below, which focuses on ranks.

Looking at specific aspects, those involving family (well-being [rank 1], happiness [2], and relationship quality [5]); health (general [3] and mental [7]); security (financial [6], about life and the future [8], and physical [21]); values (morality [4] and meaning [10]); and options (freedom of choice [9] and resources [12]) are conspicuous in their predominance at the top of the table, along with—reassuringly—some measures of happiness and life satisfaction (these are discussed below). In contrast, at the very bottom of the table, we find all four *you-only* aspects that involve relative position—power over other people [108c], social status [108b], high relative income [108a], and postmortem fame [106a]—with coefficients that are either negative or close to zero.²² Since much evidence seems to imply a *high* marginal utility to status

²¹This ranking scheme facilitates comparing the 108 *you*-aspect ranks with their corresponding 108 *everyone*-aspect ranks in the “Policy” panel, accommodating the fact that the *you-only*-aspects do not have counterparts in policy scenarios (and, similarly, that the *public*-aspects do not have counterparts in the “Personal” panel).

²²The fifth *you-only*-aspect is non-positional: “the happiness of your friends” [74a]. At the same time, two of the other aspects at the bottom—“your enjoyment of winning, competing, and facing challenges” [108] and, to a

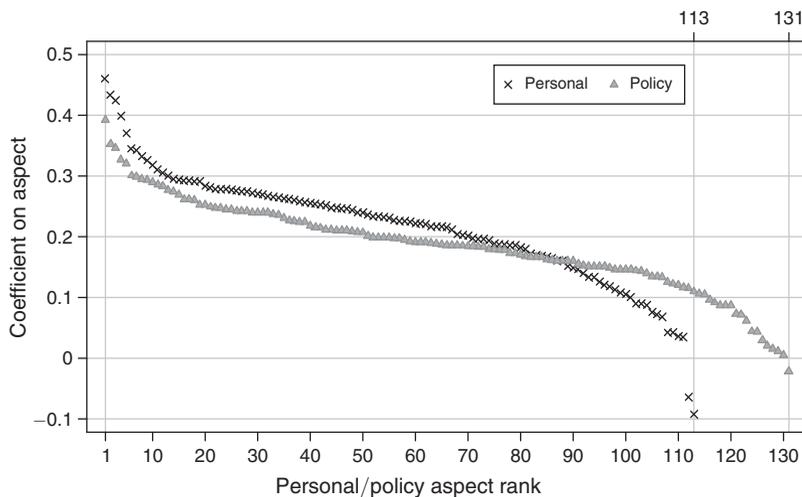


FIGURE 2. RELATIVE MARGINAL UTILITY ESTIMATES

Notes: Aspect coefficients by rank, from benchmark OLS regressions (Table 2), separately for 113 personal-scenario aspects (x's) and 131 policy-scenario aspects (triangles).

and relative position (see Heffetz and Frank 2011, for a survey), we conjecture that the low ranks of these aspects may reflect respondents' answering our stated-preference question in terms of their meta-preferences or laundered preferences.²³

While as noted above we do not view the potential for meta- or laundered preference elicitation as necessarily a disadvantage, our estimates may also be sensitive to specific details of our survey design and the underlying respondent population. Later in this section, the next section, and the online Appendix, we compare estimates based on alternative design details and on alternative subpopulations. Since our general method is held fixed, however, we can only speculate on its effect on our results. The deliberative frame of mind induced by our setup may make evaluative SWB aspects easier for respondents to consider than affective aspects; the double-negative framing of negative emotions may make them harder to think about than positive emotions; and the instruction to hold other aspects constant is almost certainly more difficult to follow when broader, combination aspects are varied in a scenario than when only narrow, fundamental aspects are varied. To the extent that our fixed design choices amplify estimated coefficients on some categories of aspects relative to others, comparisons of aspects *within* an aspect category may be more generalizable to alternative design choices than comparisons *across* categories. These caveats should be borne in mind throughout our discussion below.

lesser extent, a "ladder" aspect modeled after Cantril's Self-Anchoring Scale [103]—could have been interpreted by respondents as involving relative position, although we did not perceive them that way when compiling our aspect list. Indeed, we expected the ladder aspect to rank high, along with other evaluative SWB measures (see Section IVC below).

²³ Another possibility is that the low rank of these aspects reflects experimenter-demand effects. While we cannot rule out this concern, we believe it is less likely in an anonymous web survey like ours.

C. Personal Choices: Discussion

Evaluative and Affective SWB.—Among happiness and life satisfaction measures, the more evaluative ones—family happiness²⁴ [2] and life satisfaction [11]—are among the highest-ranking aspects, and rank higher than the more affective ones—“how much of the time you feel happy” [31] and “how happy you feel” [39]. Other measures that past work has classified as positive affect measures, such as “how often you smile or laugh” [35] (e.g., Kahneman and Deaton 2010), rank similarly to these affective happiness measures.

Negative Emotions.—Recently, Deaton et al. (2011) suggest that national SWB surveys focus also on measuring negative emotions. In our data, the six measures they recommend (in their “rough order of preference”) get the following ranks: pain [49], stress [45], worry [52], anger [76], tired (not on our list, but we have: feeling full of energy [42] and quality of sleep [77]), and sad [64]. This group of measures lies in the middle of our table, with coefficients in the range 0.19–0.25. Other negative emotions, such as frustration [67], are also in this range. Of particular interest because it is the only negative emotion among the four UK questions from the introduction, anxious [92] lies somewhat *below* this range; its coefficient (0.13) is roughly half the coefficients of stress (0.25) and pain (0.24).

Eudaimonic SWB.—While evaluative and affective measures dominate the policy discourse on national SWB surveys, researchers increasingly recognize the importance of eudaimonic dimensions of well-being. In our data, eudaimonic aspects such as being a good, moral person and living according to personal values [4] (coefficient 0.40) and having a life that is meaningful and has value [10] (coefficient 0.32) indeed rank among the highest aspects. The aspect modeled after the only eudaimonic question in the UK four—feeling that the things you do in your life are worthwhile [20]—has a reasonably high coefficient (0.28), yet lower than the two above.

In summary, our agnostic, stated-preference-based approach yields high marginal utility estimates on measures—such as life satisfaction—that have been at the center of the discussion about well-being indices. At the same time, other measures that have received recent attention—such as those of positive and negative affect, and certain eudaimonic measures—have coefficients that are not as high as aspects that have received less attention in this context, in particular, about family, health, security, values, and options. The two measures with the largest coefficients—“the overall well-being of you and your family” and “the happiness of your family”—are, as far as we know, survey questions we invented that have not previously been

²⁴ If our respondents interpret our family happiness aspect [rank 2] as including self, then it is more inclusive than typical happiness measures (that refer only to self). As a result, we do not know if its place at the top of the table is due to it being a relatively evaluative happiness measure or to its inclusion of family. (We view this aspect as a relatively evaluative measure since reporting one’s family happiness is likely to require more evaluative effort than reporting about one’s own emotions.)

asked in large-scale surveys. Our results suggest they deserve attention in future data collection efforts.²⁵

D. Policy Choices

The “Policy” panel of Table 2 reports estimates from a specification identical to that used in the “Personal” panel but uses data from the five policy scenarios each respondent faced. Recall that in such scenarios, respondents vote on policy, trading off 131 aspects that include *everyone*-aspects (personal aspects that pertain to everyone in the nation) and *public*-aspects (public goods that pertain to the entire nation or, when stated, to the entire world). To make the coefficient magnitudes comparable across the two panels, we use the same numerical scales as in the personal regression (rather than reestimating them). As mentioned above, to further facilitate such comparison between the *you*-aspects and their corresponding *everyone*-aspects, the “Rank” column ranks the *everyone*-aspects by coefficient size (1–108) and, in addition, places the 23 *public*-aspects relative to these by assigning to them rank numbers with letter suffixes.

Since we collected less data in these policy scenarios than in personal scenarios, standard errors are larger, typically in the 0.023–0.035 range. Nonetheless, the correlation between the 108 *you*- and *everyone*-coefficient pairs is fairly high (0.81).²⁶ Figure 3 conveys the comparison graphically by replicating Figure 2 for only the 108 *you*- and 108 *everyone*-aspects, both sorted by the rank of the *you*-aspects in the personal scenarios. The dashed curve reports a locally weighted linear regression of the *everyone*-coefficients (triangles). The figure suggests that on average, *everyone*-coefficients in the policy scenarios are attenuated versions of their counterpart *you*-coefficients in the personal scenarios. This may reflect respondents’ greater uncertainty regarding others’ preferences, perhaps causing respondents to state preferences with weaker intensity in policy scenarios (though remember that coefficient ratios are what matter for the index).

Consistent with the high correlation, some of the high-ranking *you*-aspects retain their high rank as *everyone*-aspects. These include overall well-being of you and your family [personal rank 1; policy rank 3], health [3; 6], personal values [4; 2], and financial security [6; 8]. At the same time, aspects related to freedom and to avoiding abuse seem to rank higher as policy aspects. These include the freedom to choose [9; 1]; your ability to pursue your dreams [36; 14]; being treated with dignity [26; 11]; and, among double negatives, avoiding deception [23; 5], pain [49; 10], and emotional abuse [68; 16].

Perhaps most importantly, several of the 23 *public*-aspects (included in the policy but not the personal scenarios) have among the largest coefficients. These include freedom from corruption, injustice, and abuse of power [0a] (coefficient 0.39), society helping those who struggle [5a], the morality of other people [5b], freedom of speech and of political participation [6a], and the well-being of the people in your

²⁵For example, to capture our overall top-ranked aspect, surveys could ask: “On a scale from zero to ten, how would you rate the overall well-being of you and your family?”

²⁶Here and later, we report the correlation between two vectors of estimated (sample) coefficients. Due to sampling error, this reported correlation is a lower bound on the correlation between the vectors of true (population) coefficients.

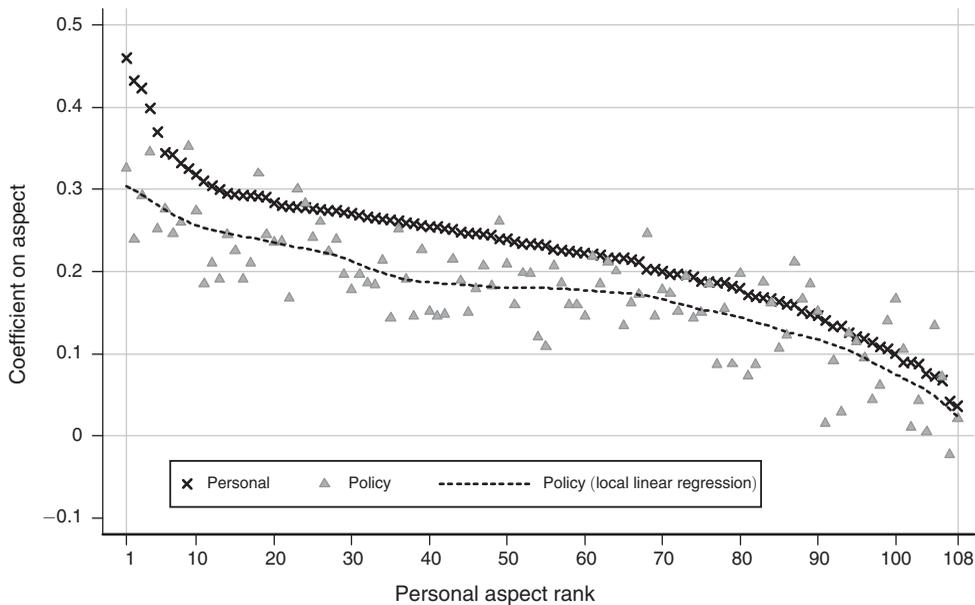


FIGURE 3. RELATIVE MARGINAL UTILITY ESTIMATES: 108 *YOU*- AND *EVERYONE*-ASPECTS

Notes: Aspect coefficients from benchmark OLS regressions (Table 2) for the 108 *you*-aspects (from personal scenarios, x's) and their corresponding 108 *everyone*-aspects (from policy scenarios, triangles), by rank in the personal scenarios. Dashed curve: local linear regression of *everyone*-coefficients (Epanechnikov kernel, bandwidth = 6).

nation [6b].²⁷ High coefficients on freedom from corruption and freedom of speech seem consistent with the tendency for respondents to weight heavily, in a policy context, aspects that expand individuals' choice sets—reducing the need to choose specific outcomes for others. We interpret these findings as providing empirical evidence that when making a policy choice, our respondents put high value on capabilities (Sen 1985) and basic rights (Rawls 1971).

Finally, we discuss “objective” aspects, some of which are modeled after widely used indicators. As discussed above, some of these aspects have large coefficients, including freedom from corruption, injustice, and abuse of power; financial security; health; and freedom of speech and of political participation. But among standard macroeconomic indicators, only low unemployment [18a] has a relatively high rank. Others have lower ranks: low inflation [43a], GDP growth [57a], GDP per capita [61a], and GDP [88a]. The tendency to prefer aspects that increase the choice set (rather than focusing exclusively on outcomes) may help explain why, among the “objective” aspects, the coefficient on equality of opportunity [19b] is relatively large (0.24), larger, for example, than that on equality of income [66a] (0.16).

²⁷Note that the *public*-aspect “the well-being of people in your nation” [6b] is quite similar to the *everyone*-aspect “the well-being of people and their families” [3]. The higher rank of the latter is consistent with the idea that highlighting “families” may raise the perceived importance of an aspect.

E. Cross-Group Heterogeneity

To provide some evidence on cross-group differences in marginal utilities, we produced versions of Table 2 for subpopulations of our respondents. Online Appendix Tables WA2–WA6 report coefficients, respectively, for men and women; those above and below an income of \$50,000/year; liberals, moderates, and conservatives; those who do and do not attend religious services at least monthly; and those younger and older than 45. The tables allow for many comparisons and tests; here we briefly summarize only some of the main findings that emerge.

Overall, aspects in the personal scenarios rank similarly across the subpopulations we examine. The sets of 113 coefficients are highly correlated across pairs of disjoint groups, with correlations ranging from 0.86 (liberals versus moderates) to 0.91 (men versus women). Moreover, there seems to be a broad consensus across the subpopulations that the highest-ranking aspects include those related to family, health, and security.

To explore which aspects' coefficients change most dramatically across groups, we first normalize our marginal utility estimates to comparable units across groups by dividing each group's set of 113 coefficients by its mean (effectively treating the "average aspect" as numeraire); we then examine, for each aspect and each pair of disjoint groups, the ratio of the two normalized coefficients, dividing the smaller coefficient by the larger (so all ratios are ≤ 1). Tables WA2–WA6 report these ratios. We highlight here aspects that rank in the top ten in one group but drop sufficiently in the complementary group to yield a ratio of less than 0.8. We caution that not only our estimates in Table 2, but also these cross-group differences, may not generalize to a more representative sample (cf. Heffetz and Rabin 2013).

Men rank higher: "your sense that your life is meaningful and has value" [men rank 5, women rank 29, normalized coefficient ratio (cr) = 0.77]. **Women rank higher:** "your mental health and emotional stability" [women 6, men 23, cr = 0.71].

High-income rank higher: life satisfaction [high income 5, low income 41, cr = 0.70], and having a meaningful life [high 7, low 24, cr = 0.75].

Liberals rank higher: having enough time and money [liberals 4, conservatives 30, cr = 0.78]. **Conservatives rank higher:** being a good, moral person and living according to personal values [co. 4, li. 6, cr = 0.75], family happiness [co. 1, li. 3, cr = 0.75], and family relationships [co. 5, li. 9, cr = 0.77].

More religious rank higher: sense of purpose [more religious 9, less religious 72, cr = 0.60], having people around you who share your values, beliefs, and interests [more 10, less 60, cr = 0.69], feeling grateful [more 8, less 50, cr = 0.74], being good and moral [more 2, less 7, cr = 0.74], and making a difference and making the world a better place [more 7, less 35, cr = 0.76]. **Less religious rank higher:** life satisfaction [less 9, more 36, cr = 0.77].

Older rank higher: having many options and possibilities in life and the freedom to choose among them [older 5, younger 35, cr = 0.65].

In policy scenarios, correlations between pairs of sets of the 131 coefficients are lower—although still reasonably high—and range from 0.60 (liberals versus conservatives) to 0.81 (more versus less religious). Using the same method as above, all ratios are reported in Tables WA2–WA6. Here we highlight aspects with normalized ratios below 0.6. **Women rank higher:** “people being good, moral people ...” [women 1, men 35, $cr = 0.50$]. **Men rank higher:** “people getting the rewards and punishments they deserve” [men 5, women 75, $cr = 0.58$]. **Liberals rank higher:** the condition of animals, nature, and the environment [li. 1, co. 113, $cr = 0.27$]. **Conservatives rank higher:** “people’s ability to have and raise children” [co. 8, li. 111, $cr = 0.33$]. **Older rank higher:** being treated with dignity and respect [older 6, younger 76, $cr = 0.53$].

V. Robustness and Additional Results

In the previous section we reported results: from (i) a simple OLS specification; including (ii) our entire personal and policy aspect lists; and pooling responses (iii) across scenarios with different numbers of aspects, (iv) across all but the speediest-to-answer respondents, and (v) across scenarios faced earlier and later in the survey. In this section, we briefly revisit these points, summarizing analyses reported in more detail in the online Appendix.

(i) *Econometric Specification.*—In our OLS specification we coded the verbal scales of the independent and dependent variables as exogenously imposed numeric scales. These were estimated from the following nonlinear ordered probit specification:

$$(4) \quad StatedPreference_s^* = (\mathbf{Option}_s \circ [\gamma_{slightly} \mathbf{1}_{\{slightly\},s} + \gamma_{somewhat} \mathbf{1}_{\{somewhat\},s} + \mathbf{1}_{\{much\},s}])' \cdot \beta + \varepsilon_s.$$

$StatedPreference_s^*$ is the latent dependent variable; \mathbf{Option}_s , $\mathbf{1}_{\{slightly\},s}$, $\mathbf{1}_{\{somewhat\},s}$, and $\mathbf{1}_{\{much\},s}$ are vectors (whose length is the number of aspects) that jointly encode the differences between the two options; and \circ is the entry-wise vector product. Each entry of \mathbf{Option}_s is equal to -1 , $+1$, or 0 , depending on whether the aspect table rates that aspect higher on Option 1, Option 2, or neither. The entries of $\mathbf{1}_{\{slightly\},s}$, $\mathbf{1}_{\{somewhat\},s}$, and $\mathbf{1}_{\{much\},s}$ are indicators of whether the aspect is rated slightly, somewhat, or much higher. We assume that ε_s is normally distributed and use maximum likelihood to estimate the parameters: $\gamma_{slightly}$, $\gamma_{somewhat}$, the coefficient vector β , and the five cutpoints that link $StatedPreference_s^*$ to the observed choice, $StatedPreference_s$. Results are reported in online Appendix Table WA7.

Note that specification (4) normalizes “much” to be -1 or $+1$ depending on whether it favors Option 1 or Option 2, and the numerical values for “slightly” and “somewhat” used for the OLS specification in Section IV are $\hat{\gamma}_{slightly}$ and $\hat{\gamma}_{somewhat}$. The resulting aspects-rating scale (i.e., the values of $\hat{\gamma}_{slightly}$ and $\hat{\gamma}_{somewhat}$) is determined by the extent to which the verbal labels affect choice differently, which in turn depends on a combination of three factors: (i) respondents’ quantitative interpretations of the

verbal labels (namely, the magnitudes on the x -axis that correspond to “slightly” and “somewhat,” given that “much” is normalized to 1); (ii) higher derivatives of utility, averaged across the aspects (i.e., how utility on the y -axis depends on the x -axis magnitudes); and (iii) any “focusing effect,” that is, respondents paying less attention to an aspect’s rating than to its direction in favor of one of the options.

To obtain values for the choice scale in the OLS specification, we use the standard normal cdf to calculate the expected value of latent preference intensity conditional on observed preference-intensity category; linearly rescale these conditional expectations to lie in the $(-1, +1)$ interval; and symmetrize them around zero by taking the average of the absolute value of each pair of corresponding conditional expectations. The resulting choice scale captures respondents’ quantitative interpretations of the verbal choice labels.

Not surprisingly, since the numerical scales in the OLS regressions reported in Section IV are estimated from the nonlinear ordered probit, the correlations between the 113 personal β ’s across Tables 2 and WA7, as well as between the 131 policy β ’s across the tables, are virtually 1. As an alternative, reestimating the OLS regressions in Table 2 using linear scales as described in Section IVB yields personal and policy coefficients (Table WA8) whose correlations with those in Table 2 are above 0.99. As a variant that allows respondents to differ in their interpretation of the choice intensities, we first normalize the choice scale at the respondent level by stretching the linear scale so that the variance across each respondent’s 30 choices is 1, and only then estimate the OLS (Table WA8). The correlations between the coefficients estimated with and without this normalization are at least 0.98. For probit and logit models, see Table WA9.

(ii) *Fundamental and Combination Aspects.*—If a combination and a fundamental aspect appear in the same scenario, then the presence of the combination aspect might affect respondents’ interpretation of the fundamental aspect, or vice-versa. For example, if “the overall well-being of you and your family” is a function of “your family’s happiness,” then a respondent asked to trade them off might—contrary to our intention—interpret the former as meaning overall well-being exclusive of family happiness. Depending on the prevalence of such situations and on how respondents interpret them, our estimated coefficients might be biased.

To probe the robustness of our results to this potential concern, we reestimate our benchmark OLS model leaving out scenarios containing aspects that seem most likely to be functions of other aspects on our list. For example, for personal choices we exclude all aspects that we view as evaluative SWB measures—including both commonly used ones (e.g., life satisfaction) and our novel proposals (e.g., how desirable your life is).²⁸ The estimated coefficients on the 94 remaining aspects (Table WA10) are broadly similar to those reported in Table 2 (correlation 0.99). Results are similar in other specifications, for example, excluding the macroeconomic indicators in policy vote scenarios (Table WA11).

²⁸ As our data and code are available on the journal’s and our websites, an interested reader can readily reestimate our model using her or his preferred subset of aspects.

(iii) *Number of Aspects per Scenario.*—As explained in Section IIIA, respondents face two-, three-, four-, and six-aspect scenarios. Reestimating our OLS model separately for each of these four scenario designs, we find that the range of coefficient sizes roughly halves from two- to six-aspect scenarios. The correlations between the four sets of 113 and of 131 coefficients range from 0.84 to 0.92 in personal and from 0.52 to 0.70 in policy scenarios (but note that the latter are more attenuated because the policy coefficients' estimation errors are larger). We also examine, for each pair of coefficient sets, the intercept of the SD line (the line going through the mean of the sets whose slope is the ratio of standard deviations). If one set were identical to another up to a multiplicative scalar—i.e., the two sets implied identical marginal rates of substitution—the intercept would be zero. We find that the intercepts are close to zero, ranging from -0.04 to 0.15 in personal and from -0.10 to 0.10 in policy scenarios. Our general conclusion is that while respondents allow each aspect to influence their stated preference less intensely *per aspect* when the number of aspects is larger, relative coefficient size remains rather stable in personal scenarios and somewhat stable in policy scenarios. Further details are provided in Table WA12 and in Figure WA1.

(iv) *Respondents' Effort and Comprehension.*—Respondents may exert little effort on unincentivized surveys. This may bias the estimated coefficients away from the true marginal utilities, for example by “compressing” aspects' coefficients toward each other if respondents pay less attention to each aspect's identity, or by biasing coefficients toward zero if respondents answer randomly. Using amount of time to complete the survey as a proxy for effort level, and reestimating our main OLS specification separately by approximate sextiles (Table WA13), we indeed find that the coefficients of the speediest sextile (less than eight minutes) are severely attenuated relative to other sextiles' coefficients. Outside the speediest sextile, coefficient sizes seem to increase with completion-time sextile less steeply and then peak at the second-slowest sextile (21–31 minutes). Furthermore, the correlations between the personal coefficients of the speediest sextile and of other sextiles range from 0.23 to 0.32, much lower than the correlations between pairs of other sextiles' coefficient sets, which range from 0.76 to 0.89. For these reasons, our benchmark OLS specification reported in Table 2 excludes respondents who took less than eight minutes to complete the survey.²⁹ Further excluding those who reported in the exit questionnaire that they did “not really” understand what they were asked to do (see Section IIIB above) yields virtually identical estimates (Table WA15).

(v) *Early versus Late Scenarios.*—Responses made later in the survey may be less reliable due to tiredness or boredom. Alternatively, they may be *more* reliable due to practice. Also, respondents' interpretations of particular aspects may change over the course of the survey. For example, respondents may interpret life satisfaction as

²⁹ As an additional respondent-effort sensitivity check, we examine respondents who, in at least one three-aspect scenario where one option happens to rate higher on all three aspects, choose the other option. The rankings of aspects in this sample are similar to our benchmark, but the coefficients are greatly attenuated (see Table WA14). We like this sensitivity check less because it may fail to drop random responders (who choose the higher-rated option by chance) and, at the same time, may unduly drop respondents who prefer *less* of certain aspects.

a broad SWB measure early on, but as they face new examples of affective SWB measures, they may interpret it to exclude feelings.

To assess these possibilities, we estimate an augmented version of our OLS specification by including a dummy for whether a scenario appeared in the earlier half of the survey and interacting it with each aspect's rating (Table WA16). We find no evidence of systematic differences between estimates from scenarios in earlier and later halves of the survey.

VI. Pragmatics

We now return to the two practical issues we posed in the introduction: which questions should an SWB survey ask, and how should the responses be weighted? In theory, the well-being index formula, $\sum_{j=1}^J \frac{\partial u(\mathbf{w})}{\partial w_j} w_j$, provides clear solutions: ask about the levels of each of the J fundamental aspects comprising \mathbf{w} , and weight the responses by their marginal utilities. In practice, as we have emphasized throughout, both our proposed list of fundamental aspects and our marginal-utility estimates are only first-pass proofs-of-concept in need of further development.

In conjunction with making progress on those fronts, governments that wish to construct reliable well-being indices will have to overcome additional challenges that we have deferred until now. In this section we return to these challenges and outline directions for surmounting them.

A. Overlapping Questions

In our theory in Section I, preferences are defined over a vector \mathbf{w} of fundamental aspects. While further progress on extending our list of survey questions may eventually result in an exhaustive list that includes *all* the components of \mathbf{w} (as required by the theory), it would likely also further increase the incidence of conceptually overlapping questions (which the theory assumes away). Note that conceptual overlap between questions is different from empirical *covariance* (e.g., in the time series) between questions: while overlap exists when two or more questions refer, by the meaning of the language they use, to some of the same fundamental aspects (an extreme example: two questions that are perfectly synonymous), covariance between questions is a feature of the empirical joint distribution of aspect levels (an extreme example: two conceptually distinct measures A and B that are perfectly correlated, say, due to A causing B). Hence nonzero covariances are likely even in the absence of overlap, and information regarding covariances across SWB survey waves (or across respondents) is not sufficient for identifying overlap.

Intuitively, including overlapping questions in the well-being index would lead to double-counting, a problem analogous to counting some components of GDP more than once. Hence, before a reliable well-being index can be constructed, a strategy is needed for avoiding double-counting. One such strategy is to find a list of SWB questions free from overlap by eliminating questions that overlap with other questions on the list. Doing so requires a method for detecting overlap between any given pair of SWB survey questions. This subsection briefly outlines the basic idea

underlying one such method; we provide an extended discussion, a fully specified concrete example, and a formal treatment in the online Appendix.

The idea behind our overlap-detection method is to use a variation on our SP survey in order to compare the (relative) marginal utility of the *joint* increase in two survey questions (appearing within a single scenario option) with the sum of the (relative) marginal utilities of the *separate* increases of the two survey questions (appearing in two separate scenarios). For example, consider two of the UK questions from the introduction: life satisfaction and life worthwhileness. If the two questions do not overlap—that is, if they elicit two fundamental aspects or, more generally, two functions of disjoint sets of fundamental aspects—then the sum of the marginal utility of life satisfaction plus the marginal utility of life worthwhileness (both relative, say, to the marginal utility of health) would equal the (relative to health) marginal utility of a joint increase in life satisfaction and life worthwhileness. This equality holds approximately even if the aspects measured by the two questions enter preferences as complements, as long as the increases are small. On the other hand, if the two questions do overlap, then the marginal utility of their joint increase would be smaller than the sum of marginal utilities of their separate increases because part of the joint increase is overlapping and is taken into account by survey respondents only once.

Therefore, to identify overlap between, e.g., life satisfaction and life worthwhileness, in the online Appendix we propose to measure these (relative) marginal utilities by collecting data from (i) two-aspect scenarios where respondents choose between different increases in life satisfaction in Option 1 versus increases in a third aspect in Option 2; (ii) two-aspect scenarios where respondents choose between different increases in life worthwhileness in Option 1 versus increases in that third aspect in Option 2; and (iii) three-aspect scenarios where respondents choose between different increases in life satisfaction *and* life worthwhileness in Option 1 versus increases in that third aspect in Option 2. Such a data collection effort would effectively replicate our own SP survey, with three main differences: first, it would target pairs of questions where overlap is suspected; second, it would involve only the relevant two- and three-aspect scenarios described in (i), (ii), and (iii) above; and third, it would adjust and refine the language in the scenarios' preamble to encourage respondents to interpret joint increases in pairs of questions in specific ways, a point that we briefly discuss now.

While our SP-survey instructions may be sufficient when the SWB survey questions under study are fundamental aspects, the formal theory we develop in the online Appendix highlights the importance of respondents' interpretation of the implied changes in fundamental aspects when the questions under study are not themselves one-to-one measures of fundamental aspects. We present two propositions, relying on two alternative sets of assumptions regarding respondents' interpretations, each of which formally substantiates the validity of our overlap-detection method. However, as we discuss in the online Appendix, developing the survey language that would encourage respondents to interpret scenarios in one of the ways required by our theory would necessitate a process of developing and testing survey instructions that goes beyond the scope of this paper. Nevertheless, our Appendix treatment suggests that the problem of potential question overlap can be analyzed formally and addressed practically, and takes some tentative steps in those directions.

B. Abridged Index

From a pragmatic point of view, identifying a sub-list of survey questions that avoids overlap (while fully covering the fundamental aspects in \mathbf{w}) has an additional important consequence: it shortens the list of questions to elicit. In what follows, we use the term “full index” to refer to the well-being index that is based on such a shorter yet complete list, which we assume consists of exactly J fundamental-aspect questions.³⁰ Such a list may still however be longer than the list of questions a government is able (or willing) to regularly include on each wave of its SWB survey. It may hence be important to also consider the two practical questions from the introduction with an added constraint: which items should a government include on an SWB survey that is limited to only $N < J$ questions? And, conditional on asking N questions (not necessarily the optimal ones), how should the responses be weighted?

Conditional on an SWB survey eliciting responses $\{r_1, r_2, \dots, r_N\}$ to a given set of N questions, a natural weighting approach is to seek weights, $\alpha_1, \alpha_2, \dots, \alpha_N$, such that the “abridged index” $\sum_{n=1}^N \alpha_n r_n$ is the best predictor of the full index in an R^2 sense. The optimal weights will generally no longer be proportional to the estimated marginal utilities because included questions will proxy for excluded aspects with which they covary in the time series.³¹ If, in addition, the N questions can be selected, then a natural selection approach is to select those that, weighted optimally, best predict the time series of the full index. (Combination-aspect questions that have been excluded from the full index due to overlap may be especially useful in this context.) Note that the optimal set of N questions may not necessarily include the questions with the largest marginal-utility estimates. For example, Deaton et al. (2011) argue that negative emotions (whose coefficients we report in Section IVC above as ranking only in the middle of our Table 2) are important to include on a government’s SWB survey due to their relatively high variance and low covariance with other questions.

To construct an N -question abridged index that is optimal in the above sense of best-predicting the full index, it would be necessary to first construct and track the full index, at least for a few survey waves. In the theoretically ideal world of an infinite sample and unlimited computing power, the optimal abridged index would be found by regressing the full index on all possible sets of N questions, finding the set that maximizes R^2 . In practice, with a finite sample, more sophisticated methods are needed to avoid overfitting in both selection of questions and estimation of their coefficients. Miller (2002) provides a comprehensive discussion of such methods.

Two additional points regarding an abridged index are worth mentioning. First, different abridged indices may be optimal for different intended uses of the index. For example, if the abridged index is used mainly to track individuals’ well-being

³⁰With J fundamental aspects in \mathbf{w} , a complete no-overlap list has at most J questions: exactly J if they are all fundamental-aspect questions, or fewer if some are “composite-aspect” questions, meaning that they elicit non-fundamental aspects that are subutility functions (see the online Appendix for a formal definition and discussion). Such composite aspects can be substituted in the theory and in the index for their underlying fundamental aspects.

³¹Note that variants of our SP survey (including those referred to in Section VIA) are not informative regarding covariances because the SP survey exogenously varies the hypothetical aspect bundles to be compared in a way unrelated to time-series covariances of responses to the SWB survey.

over time, then the objective function to best-predict is a time-series of the full index (as proposed above). Alternatively, if the index is used to guide policy, then question selection and weighting should account for the fact that some aspects may vary a great deal but be relatively immune to policy, while others may move little unless changed by policy (for example, Deaton et al. 2011 argue that stress may be particularly sensitive to policy).

Second, if the questions' variance-covariance matrix shifts over time, then an abridged index's optimal questions and weights may change even when the intended use is held fixed and even if the marginal utilities have not changed. This point is related to the well-known Lucas critique from macroeconomics; even if the underlying structural model is fixed, the best-fitting reduced-form equation may be unstable as circumstances shift. For this reason, the full index should be tracked at least periodically—e.g., by switching back, every few survey waves, from the abridged survey to the full survey—and the optimal abridged index should be periodically reestimated.

C. Pooling Respondents

While abridged indices would reduce the number of questions to ask on regular SWB-survey waves, our discussion above makes it clear that at least periodically, it is crucial to construct the full index. Hence, at least periodically, the full list of SWB questions *and* a full set of marginal-utility estimates would be needed. Even if the full list of SWB questions is not by itself considered too long to ask on a single SWB survey, multiple SP-survey questions are required for estimating each marginal utility, and therefore estimating the full set of marginal utilities may require a long survey. While breaking a long survey into shorter modules administered to the same respondent on successive occasions may be feasible, a government might wish to divide the full set of required SP questions across different respondents and pool their responses. Under what conditions would doing so be justified?

Our theory in Section I is restricted to an individual agent. Estimating marginal utilities from pooled data is therefore theoretically justified only when the pooled respondents' indifference surfaces have the same local slope. We refer to such respondents as being of the same "type."³² In practice, while it is unlikely that any two respondents are of exactly the same type, it may be possible to partition a population of respondents into approximate types. To do so, one would need to conduct a full SP survey (long enough for identifying each respondent's local marginal utilities) on at least a subsample of respondents. One could then search for observables (such as sex, age, etc.) that may characterize types.

In our own main specification in Section IV, we pooled data across all of our respondents, effectively treating them all as a single type. Doing so is difficult to justify theoretically, and as discussed above governments should not—and would

³²Thus, even if two respondents have the same preferences, if their interpretation and reporting of their w 's differ in a way that leads to indifference surfaces with different local slopes, then the respondents should be considered of different types. On the other hand, even if two respondents interpret an aspect differently—for example, "being a good, moral person" could involve different behaviors that satisfy different moral values for different people—the two are still of the same type as long as that aspect's relative marginal utility is the same at their current fundamental-aspect levels.

not have to—do so. That said, to the extent that some of our estimates by several different demographic groups (reported in the online Appendix and summarized in Section IVE) are viewed as not that different from each other, it is possible that, as an empirical matter, pooling may yield sensible estimates.

VII. Concluding Remarks

Our current system of national accounts has been continually refined over many decades, and continues to be revised. Converging on how well-being should be tracked is likely to be similarly arduous. Much work has already been done. But that work, and our own contributions—the theory and methodology, aspect list, and marginal-utility estimates—are only a bare beginning. We address only a fraction of the issues in constructing well-being indices that we are aware of, and there are surely many other issues we have not thought of.

Perhaps the most urgent unresolved theoretical issue is aggregation across individuals. This issue is not specific to our approach. As a workaround, researchers (say, using aggregate consumption or GDP as a welfare measure) often take the leap of assuming a representative agent. If one were willing to assume a representative agent, a national well-being index could be constructed from marginal-utility estimates like ours, together with average responses to an SWB survey. Alternatively, including money in the SP survey as if it were an “aspect” would enable scaling the marginal utilities in dollar units (see footnote 14). The change in the well-being index could then be measured in dollars for each respondent and summed across respondents. As another alternative, the same procedure could be followed with some other numeraire good; for example, time might be an attractive numeraire. Other recent proposals regarding aggregation in the context of SWB surveys include Fleurbaey, Schokkaert, and Decancq (2009); Fleurbaey and Maniquet (2011); and Benjamin, Heffetz, Kimball, and Szembrot (2013).

Another set of concerns, emphasized by Frey and Stutzer (2007, 2012), hereafter FS, is that if aggregated SWB responses were regularly used for policymaking, both governments and individuals might have an incentive to manipulate the survey-based index for their own benefit. One concern FS raise is that politicians and public officials have many degrees of freedom that they may exploit when selecting questions, weights, and the respondent population used for constructing an index. As FS observe, similar concerns also arise with traditional indicators such as GDP and the rate of unemployment. We believe that—as with these traditional and widely-used indicators—having standardized procedures for constructing the index, such as the procedures developed in this paper, can eliminate many of these degrees of freedom. Another concern FS raise is that individuals may have an incentive to deviate from truthfully responding to the SWB survey. While we are not aware of evidence of such conscious attempts by individuals or interest groups to manipulate existing survey-based indicators such as the rate of unemployment, we explore in a companion paper (Benjamin, Heffetz, Kimball, and Szembrot 2013) a mechanism for aggregating SWB responses and guiding policy that reduces incentives for non-truthful responding.

Yet another set of concerns relates to measurement. While we have assumed throughout that aspects of well-being can be meaningfully measured with an SWB survey,

such measurement still faces major challenges that we have not addressed. For example, traditional SWB measures may be over-sensitive to immediate context, under-sensitive to lasting changes in life circumstances, and subject to recalibration of the response scale (see Adler 2013, for a comprehensive critical review of these and other challenges). And of course *all* survey measures are subject to measurement error.

We have focused on constructing a well-being index based on combining an SWB survey with an SP survey. But the framework and method we have developed could be applied in three additional directions. First, as mentioned above, one could use an SP survey to obtain weights for existing indices of objective measures, such as the Human Development Index (HDI), that currently use ad hoc weights. In that case, the objective measures—for the HDI: longevity, education, and GDP per capita—would replace the aspects in the SP survey.

Second, one could extend market-price-based indices, such as GDP, to incorporate other factors using “price” imputations. Rather than a function of fundamental aspects, utility would be modeled as a function of market goods as well as nonmarket goods such as leisure, social relationships, and the environment. These goods would replace the aspects in the SP survey.

Third, while our SP survey and analysis are based on the assumption that respondents’ stated preferences can be used to assess welfare—an approach we find attractive on liberalist grounds—our methodology could be adapted to accommodate alternative assumptions. For example, if one assumed that life satisfaction equaled welfare, then one would replace our stated choice question with a predicted-life-satisfaction question. Alternatively, one might replace “you” with “someone like you” in the stated-choice question if one believed that the latter would yield more reliable responses.

While we view this paper primarily as proposing a long-term agenda, our findings also point to a few readily actionable steps. First, our results suggest prioritizing the measurement of aspects related to family, health, and security; eudaimonic and especially evaluative SWB measures; and, especially in the policy context, freedoms and capabilities. Second, as discussed in Section VI, for the purpose of selecting specific questions for regular inclusion on SWB surveys, we highlight the value of gathering data on as many aspects as possible—at least initially (cf. Deaton et al. 2011) and at regular intervals. Third, along with conducting SWB surveys, we call for governments and researchers to devote resources to estimating aspects’ marginal utilities; our SP survey illustrates one method for doing so.

REFERENCES

- Adler, Matthew D. 2013. “Happiness Surveys and Public Policy: What’s the Use?” *Duke Law Journal* 62 (8): 1509–1601.
- Adler, Matthew D., and Paul Dolan. 2008. “Introducing a ‘Different Lives’ Approach to the Valuation of Health and Well-Being.” University of Pennsylvania Law School Institute for Law and Economics Research Paper 08-05.
- Alkire, Sabina. 2002. “Dimensions of Human Development.” *World Development* 30 (2): 181–205.
- 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. 2013. “Can Marginal Rates of Substitution Be Inferred from Happiness Data? Evidence from Residency Choices.” National Bureau of Economic Research Working Paper 18927.

- Benjamin, Daniel J., Ori Heffetz, Miles S. Kimball, and Nichole Szembrot.** 2013. "Aggregating Local Preferences to Guide Marginal Policy Adjustments." *American Economic Review* 103 (3): 605–10.
- Benjamin, Daniel J., Ori Heffetz, Miles S. Kimball, and Nichole Szembrot.** 2014. "Beyond Happiness and Satisfaction: Toward Well-Being Indices Based on Stated Preference: Dataset." *American Economic Review*. <http://dx.doi.org/10.1257/aer.104.9.2698>.
- Camerer, Colin F., and Robin M. Hogarth.** 1999. "The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework." *Journal of Risk and Uncertainty* 19 (1–3): 7–42.
- Camerer, Colin F., Samuel Issacharoff, George Loewenstein, Ted O'Donoghue, and Matthew Rabin.** 2003. "Regulation for Conservatives: Behavioral Economics and the Case for 'Asymmetric Paternalism.'" *University of Pennsylvania Law Review* 151 (3): 1211–54.
- Deaton, Angus, Daniel Kahneman, Alan Krueger, David Schkade, Norbert Schwarz, and Arthur Stone.** 2011. Memo to the Office of National Statistics' Advisory Group on Subjective Well-Being. *Supporting Documents for the Meeting to Provide Guidance to the Organisation for Economic Co-operation and Development on its Plans to Measure Self-Reported Well-Being*. Paris: Organisation for Economic Co-operation and Development (OECD).
- Decancq, Koen, and Maria Ana Lugo.** 2013. "Weights in Multidimensional Indices of Well-Being: An Overview." *Econometric Reviews* 32 (1): 7–34.
- Decancq, Koen, Luc Van Ootegem, and Elsy Verhofstadt.** 2011. "What If We Voted on the Weights of a Multidimensional Well-Being Index? An Illustration with Flemish Data." University of Antwerp Center for Social Policy (CSB) Working Paper 11/10.
- Diener, Ed.** 2000. "Subjective Well-Being: The Science of Happiness and a Proposal for a National Index." *American Psychologist* 55 (1): 34–43.
- Diener, Ed.** 2006. "Guidelines for National Indicators of Subjective Well-Being and Ill-Being." *Applied Research in Quality of Life* 1 (2): 151–57.
- Diener, Ed, and Martin E. P. Seligman.** 2004. "Beyond Money: Toward an Economy of Well-Being." *Psychological Science in the Public Interest* 5 (1): 1–31.
- Ding, Min, Rajdeep Grewal, and John Liechty.** 2005. "Incentive-Aligned Conjoint Analysis." *Journal of Marketing Research* 42 (1): 67–82.
- Dolan, Paul, Richard Layard, and Robert Metcalfe.** 2011. *Measuring Subjective Well-Being for Public Policy*. Newport, South Wales: Office for National Statistics.
- Fleurbaey, Marc.** 2009. "Beyond GDP: The Quest for a Measure of Social Welfare." *Journal of Economic Literature* 47 (4): 1029–75.
- Fleurbaey, Marc, and Didier Blanchet.** 2013. *Beyond GDP: Measuring Welfare and Assessing Sustainability*. Oxford: Oxford University Press.
- Fleurbaey, Marc, and Francois Maniquet.** 2011. *A Theory of Fairness and Social Welfare*. Cambridge: Cambridge University Press.
- Fleurbaey, Marc, Erik Schokkaert, and Koen Decancq.** 2009. "What Good is Happiness?" Université catholique de Louvain, Center for Operations Research and Econometrics (CORE) Discussion Paper 2009017.
- Frey, Bruno, and Alois Stutzer.** 2007. "Should National Happiness be Maximized?" Zurich Institute for Empirical Research in Economics (IEER) Working Paper 306.
- Frey, Bruno S., and Alois Stutzer.** 2012. "The Use of Happiness Research for Public Policy." *Social Choice and Welfare* 38 (4): 659–74.
- Gilbert, Daniel T.** 2006. *Stumbling on Happiness*. New York: Knopf.
- Graham, Carol.** 2011. *The Pursuit of Happiness: An Economy of Well-Being*. Washington, DC: Brookings Institution Press.
- Green, Paul E., and Vithala R. Rao.** 1971. "Conjoint Measurement for Quantifying Judgmental Data." *Journal of Marketing Research* 8 (3): 355–63.
- Heffetz, Ori, and Robert H. Frank.** 2011. "Preferences for Status: Evidence and Economic Implications." In *Handbook of Social Economics*, Vol. 1A, edited by Jess Benhabib, Alberto Bisin, and Matthew Jackson, 69–91. Amsterdam: Elsevier B.V.
- Heffetz, Ori, and Matthew Rabin.** 2013. "Conclusions Regarding Cross-Group Differences in Happiness Depend on Difficulty of Reaching Respondents." *American Economic Review* 103 (7): 3001–21.
- Kahneman, Daniel, and Angus S. Deaton.** 2010. "High Income Improves Evaluation of Life but not Emotional Well-Being." *Proceedings of the National Academy of Sciences* 107 (38): 16489–93.
- Kahneman, Daniel, and Alan B. Krueger.** 2006. "Developments in the Measurement of Subjective Well-Being." *Journal of Economic Perspectives* 20 (1): 3–24.
- Lancaster, K. J.** 1966. "A New Approach to Consumer Theory." *Journal of Political Economy* 74 (2): 132–57.

- Layard, Richard.** 2005. *Happiness: Lessons from a New Science*. New York: The Penguin Press.
- Loewenstein, George, and Peter A. Ubel.** 2008. "Hedonic Adaptation and the Role of Decision and Experience Utility in Public Policy." *Journal of Public Economics* 92 (8-9): 1795–1810.
- Maslow, Abraham.** 1946. "A Theory of Human Motivation." In *Twentieth Century Psychology: Recent Developments in Psychology*, edited by Philip Lawrence Harriman, 22–48. New York: The Philosophical Library, Inc.
- Micklewright, John.** 2001. "Should the UK Government Measure Poverty and Social Exclusion with a Composite Index?" Unpublished.
- Miller, Alan J.** 2002. *Subset Selection in Regression*. 2nd ed. Boca Raton: Chapman & Hall/CRC.
- Nussbaum, Martha C.** 2000. *Women and Human Development: The Capabilities Approach*. Cambridge, UK: Cambridge University Press.
- Office for National Statistics (ONS).** 2011. *Initial Investigation into Subjective Well-being from the Opinions Survey*. Newport, South Wales: ONS.
- Rawls, John.** 1971. *A Theory of Justice*. Cambridge, MA: Harvard University Press.
- 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 (6): 1069–81.
- Sen, Amartya.** 1979a. "Utilitarianism and Welfarism." *Journal of Philosophy* 76 (9): 463–89.
- Sen, Amartya.** 1979b. "The Welfare Basis of Real Income Comparisons: A Survey." *Journal of Economic Literature* 17 (1): 1–45.
- Sen, Amartya.** 1985. *Commodities and Capabilities*. Oxford: Oxford University Press.
- Sen, Amartya.** 1992. *Inequality Reexamined*. Cambridge, MA: Harvard University Press.
- Stiglitz, Joseph E., Amartya Sen, and Jean-Paul Fitoussi.** 2009. *Report by the Commission on the Measurement of Economic Performance and Social Progress*. Commission on the Measurement of Economic Performance and Social Progress. www.stiglitz-sen-fitoussi.fr.