

NBER WORKING PAPER SERIES

CHALLENGES IN CONSTRUCTING A SURVEY-BASED WELL-BEING INDEX

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

Working Paper 23111
<http://www.nber.org/papers/w23111>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2017

We are grateful for NIH/NIA grants R01-AG040787 to the University of Michigan, and R01-AG051903 to the University of Southern California, and to the Michigan Institute for Teaching and Research in Economics for financial support; to Alberto Bisin, Angus Deaton, Marc Fleurbaey, and Arthur Stone for helpful comments and discussion; and to Tuan Anh Viet Nguyen, Rebecca Royer, and Robbie Strom for outstanding research assistance. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

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 They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2017 by Daniel J. Benjamin, Kristen Cooper, Ori Heffetz, and Miles S. Kimball. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Challenges in Constructing a Survey-Based Well-Being Index
Daniel J. Benjamin, Kristen Cooper, Ori Heffetz, and Miles S. Kimball
NBER Working Paper No. 23111
January 2017
JEL No. C43,C83,D11,I31

ABSTRACT

Many in both government and academia are showing renewed interest in developing new measures of national well-being. A new measure that goes “beyond GDP” to comprehensively capture non-market goods could be a useful supplement to traditional economic indicators for guiding policy and more accurately tracking welfare. But how should national well-being be conceptualized in theory? How could it be measured in practice? How could it be constructed in a systematic and politically neutral way? These questions should be approached by economists with the same level of care that has been taken in the theoretical and practical development of GDP.

In this short paper, we focus on one conceptual framework (Benjamin, Heffetz, Kimball, and Szembrot, 2014), which uses self-reported responses to subjective well-being (SWB) and stated preference (SP) survey questions to construct an index of well-being. We briefly review the framework and highlight challenges in the first two steps a government agency would need to take before conducting the SWB and SP surveys: (1) formulating a set of aspects of well-being that is theoretically valid and can be measured accurately via surveys; and (2) choosing and interpreting the surveys’ response scales.

Daniel J. Benjamin
Center for Economics and Social Research
University of Southern California
635 Downie Way, Suite 312
Los Angeles, CA 90089-3332
and NBER
daniel.benjamin@gmail.com

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

Kristen Cooper
324 Frost Hall
255 Grapevine Road
Wenham, MA 01984
kristen.cooper@gordon.edu

Miles S. Kimball
Department of Economics
University of Michigan
Ann Arbor, MI 48109-1220
and NBER
mkimball@umich.edu

Challenges in Constructing a Survey-Based Well-Being Index

By DANIEL J. BENJAMIN, KRISTEN B. COOPER, ORI HEFFETZ, AND MILES KIMBALL

* Benjamin: Center for Economic and Social Research, University of Southern California, 635 Downey Way, Suite 312, Dauterive Hall, Los Angeles, CA 90089 (email: djbenjam@usc.edu). Cooper: Department of Economics and Business, Gordon College, 255 Grapevine Road, Wenham, MA 01984 (email: kristen.cooper@gordon.edu). Heffetz: Samuel Curtis Johnson Graduate School of Management, Cornell University, 324 Sage Hall, Ithaca, NY 14853 and Department of Economics, Hebrew University of Jerusalem (email: oh33@cornell.edu). Kimball: Department of Economics, University of Colorado Boulder, 256 UCB, Boulder, CO, 80309-0256 (email: miles.kimball@colorado.edu). We are grateful for NIH/NIA grants R01-AG040787 to the University of Michigan, and R01-AG051903 to the University of Southern California, and to the Michigan Institute for Teaching and Research in Economics for financial support; to Alberto Bisin, Angus Deaton, Marc Fleurbaey, and Arthur Stone for helpful comments and discussion; and to Tuan Anh Viet Nguyen, Rebecca Royer, and Robbie Strom for outstanding research assistance.

Many in both government and academia are showing renewed interest in developing new measures of national well-being. A new measure that goes “beyond GDP” to comprehensively capture non-market goods could be a useful supplement to traditional economic indicators for guiding policy and more accurately tracking welfare. But how should national well-being be conceptualized in theory? How could it be measured in practice? How could it be constructed in a systematic and politically neutral way? These questions should be approached by economists with the same level of care that has been taken

in the theoretical and practical development of GDP.

In this short paper, we focus on one conceptual framework (Benjamin, Heffetz, Kimball, and Szembrot, 2014; hereafter BHKS), which uses self-reported responses to subjective well-being (SWB) and stated preference (SP) survey questions to construct an index of well-being. We briefly review the framework and highlight challenges in the first two steps a government agency would need to take before conducting the SWB and SP surveys: (1) formulating a set of aspects of well-being that is theoretically valid and can be measured accurately via surveys; and (2) choosing and interpreting the surveys’ response scales.

We focus on constructing a personal well-being (PWB) index and do not address here the problem of interpersonal aggregation of PWB indices into a measure of national well-being. Among existing approaches to aggregation, we believe that recent research on methods that aggregate ordinal utilities is the most promising (for example, approaches building on money-metric utilities as in Fleurbaey and Blanchet, 2013).

I. Theoretical Framework

A consensus is emerging that well-being is multi-dimensional (Stiglitz, Sen, and Fitoussi, 2009), and evidence suggests that—despite earlier hopes—it is unlikely to be fully captured by a single catch-all measure such as a happiness or life-satisfaction question (e.g., Benjamin, Heffetz, Kimball, and Rees-Jones, 2012). But if the levels of each dimension of well-being can be measured, and if the relative importance of each dimension can be estimated, then an index may be constructed that could track well-being.

To this end, in BHKS we proposed a simple framework that interprets well-being as preference satisfaction and uses standard utility theory to derive a PWB index. Our approach is analogous to the theory behind the measurement of aggregate consumption. That theory starts with a utility function $u(\mathbf{c})$ defined over a consumption vector for M goods. A traditional aggregate consumption index, $\sum_{m=1}^M \bar{p}_m c_m$, weights each good's consumption, c_m , by its price held fixed at a baseline level, \bar{p}_m . In the face of small changes in consumption, changes in the index approximate changes in utility (up to a multiplicative constant): $\sum_{m=1}^M \bar{p}_m \Delta c_m \propto \sum_{m=1}^M \frac{\partial u(\mathbf{c})}{\partial c_m} \Delta c_m \approx \Delta u$.

We proposed replacing the vector of M consumption goods, \mathbf{c} , with a vector of J “fundamental aspects of well-being,” \mathbf{w} . These aspects are the final goods that people ultimately care about (i.e., consumption goods are now treated as intermediate goods in the production of aspects of well-being, à la Lancaster, 1966). The aspects may be objective (directly measurable) or subjective (self-reported), and objective measures might eventually be available for some of the now-subjective aspects (e.g., biometrics for certain health aspects and even emotional states). In the implementation we discuss here, the levels of \mathbf{w} are measured with SWB surveys, so the aspects are either inherently subjective or subjective perceptions of objective aspects. We discuss objective aspects further in section IV.

Because the aspects are not traded in markets, price data are unavailable and different individuals may have different marginal rates of substitution (MRSs) across the aspects. Nevertheless, a PWB index can be constructed using each individual's MRSs for the aspects as weights. Specifically, the index is given by $\text{PWB} \equiv \sum_{j=1}^J \frac{\partial u(\mathbf{w})}{\partial w_j} w_j$, where the marginal utilities (MUs) are defined relative to an arbitrary numeraire aspect. Small changes in this PWB index provide a first-order approximation to changes in the individual's

(ordinal) utility (even if the individual's preferences are non-linear).

Figure 1 shows examples of potential SWB questions for measuring the levels of two aspects of well-being, related to happiness and meaningfulness. Figure 2 shows an example stated preference survey question for measuring the MRS between these aspects. Both are taken from an ongoing project of ours that attempts a large-scale web survey implementation of the BHKS framework.

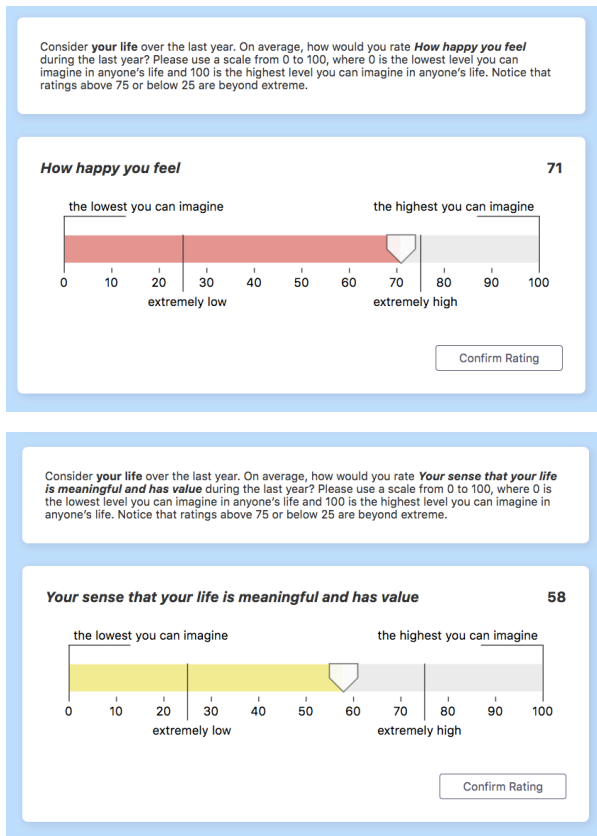


Figure 1. Two sample SWB survey questions.

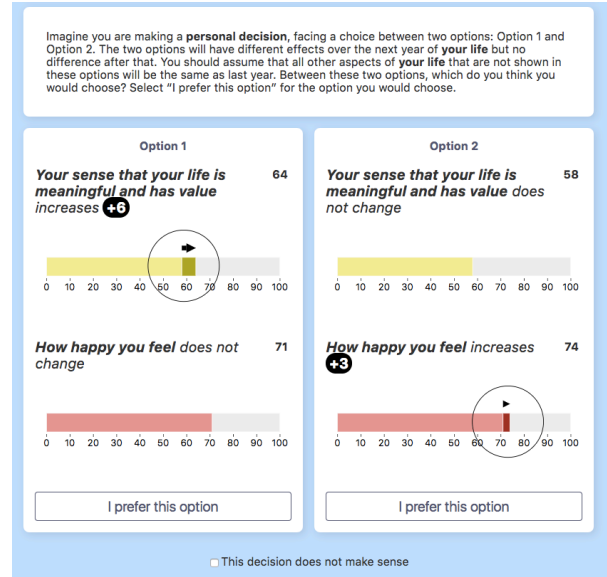


Figure 2. A sample SP survey question.

II. Challenge #1: Formulating the Set of Aspects

Within our theoretical framework, any set of J aspects of well-being can be used for the index as long as it satisfies two properties: *comprehensiveness* and *non-overlappingness*. To implement the framework with surveys, every aspect in the set must also satisfy a third, cognitive, requirement: *accessibility*.

Comprehensiveness means that the set covers all aspects of well-being that matter to the individual. This is a counterpart to the requirement that expenditure-based indices such as GDP cover all potential types of spending (e.g., both goods and services).

Non-overlappingness means that each of the aspects in the set has its own distinct contribution to preferences. For GDP, if the value of restaurant meals and the value of

alcoholic beverages were independently added up, then the value of alcohol consumed in restaurants would be double-counted. Similarly, a PWB index will suffer from a double-counting problem if an ultimate object of desire that enters preferences once is counted more than once. For example, if “how much you like your life” and “how much you enjoy your life” appear as two different aspects in the set and get equally high marginal utilities, but mean the same thing to people and enter their preferences once, then only one of these aspects should be in the set.

The solution in the case of GDP is to define the expenditure categories so that they are conceptually distinct. In BHKS we proposed, but did not test, a method of detecting conceptual overlap between aspects based on the idea that the sum of MUs for two overlapping aspects will exceed the MU of an aspect generated by concatenating the two. We hope that such a method could be applied to prune the set. If pruning cannot eliminate overlap without jeopardizing comprehensiveness, it is worth exploring ways to adjust the index for the remaining overlap.

Accessibility means that respondents can accurately introspect and report about (i) their own level of an aspect of well-being, w_j , and (ii) how it affects their welfare, $\frac{\partial u(w)}{\partial w_j}$. These requirements are analogous to the assumptions

implicit in GDP construction that (i) we have access to accurate consumption (quantity) data and (ii) market prices reflect consumers’ true MUs.

Existing research suggests that survey-based measures of well-being may violate accessibility in systematic ways, for example due to their apparent sensitivity to contextual details such as question order. The reasons for such sensitivity, its practical implications, and potential solutions are still actively researched and hotly debated (e.g., Deaton and Stone, 2016, and the comment and authors’ response in the same journal issue).

Social-desirability reporting biases may also pose a challenge to accessibility: when asked about “dirty” preferences (e.g., racism), people may consciously or unconsciously launder their responses. It is sometimes argued that such laundered preferences are more relevant for normative purposes, in which case the resulting MUs should be used in the PWB. In other cases, it may be possible to find a related aspect of well-being that is socially acceptable enough to mitigate the bias. For example, one can ask about “you feeling powerful” rather than “you having power over other people.”

To date, we have relied on introspection to assess whether aspects are accessible, but we

would like to see a more systematic examination.

The challenge of formulating the aspect set poses several tradeoffs. In BHKS, we scoured the economics, psychology, and philosophy literatures for lists of what matters to people, and we came away with over 100 aspects. Since BHKS, we have further expanded our set to over 2000 (!) potential aspects of well-being. Having such a large set likely improves comprehensiveness but exacerbates overlap. Beginning with such a relatively exhaustive set, we plan to learn through empirical testing which potential aspects have low enough MUs or are duplicative enough that little is lost by omitting them. One would wish to end up with an aspect set that is small enough that every survey respondent can answer questions about all aspects in the set on each survey occasion. Then, individual-level MRSs can be estimated and preference heterogeneity fully accommodated. A larger aspect set necessitates pooling data within groups of respondents, and such pooling is only justified theoretically if respondents within each group have homogeneous MRSs.

III. Challenge #2: Choosing the Response Scales

In the index, $\sum_{j=1}^J \frac{\partial u(w)}{\partial w_j} w_j$, the units of w_j cancel out when w_j is multiplied by its MU.

Therefore, perhaps surprisingly, in theory it does not matter what response scale we choose for an aspect—a 0-10 scale, a 0-100 scale, an amount-of-smileyface scale, whatever—as long as it is the same response scale in the SWB and SP surveys (as in Figures 1 and 2) and the respondent uses the response scale the same way in both surveys. Indeed, the theory allows for different scales for different aspects, and for different respondents to use the same scale differently.

In practice, a respondent may not always use the response scale the same way across the SWB and SP surveys. For example, if respondents are uncomfortable with numbers, they may be less attentive to the magnitudes of the tradeoffs in the SP survey than to the numerical levels in the SWB survey. This could lead to both noisier responses and systematic biases in the MRS estimates.

An individual's *shifts* in scale use over time are another concern. If a change in a reported w_j reflects a shift in scale use rather than an actual change in the aspect level, then the resulting change in the index will be misinterpreted as a change in well-being.

Systematically studying possible shifts in scale use and developing ways of correcting for them is a high priority. One approach would be to find aspects of well-being that can be measured biometrically and compare them

to survey measures over time. In another undertested approach, respondents in a panel would rate aspects of well-being not only in their own current lives but also in hypothetical situations that are held fixed across survey occasions. Finally, it is worth exploring whether changes in a respondent's scale use across two survey occasions could be detected by comparing the change in her reported aspect level with her answer to a direct question about the change.

One reason a respondent might shift scale use is to deal with the ceiling on a scale. This issue has received too little attention. Scales can be designed to reduce top-coding issues by, say, labeling the top of the scale "extremely happy" rather than "very happy." Figure 1 shows our approach of labeling the top of the scale "the highest you can imagine in anyone's life." It remains to be investigated how effective this labeling scheme is in reducing top-coding.

IV. Discussion

Given space constraints, we have focused on the issues that seem most pressing to address before governments can begin collecting data that can eventually be used for constructing theoretically valid well-being indices. Yet we are also concerned about many other issues. Here we briefly mention

three that seem especially important but that we are not as far along in thinking through.

One open question is how to decide at what *level of generality* to specify the aspects, e.g., "your health" vs. components of health. We conjecture that it matters because there may be a "part-whole bias": the sum of the MUs estimated for an aspect's components may exceed the MU estimated for the aspect considered holistically. We suspect that a reasonable rule of thumb is to try to specify aspects such that they have estimated MUs of the same order of magnitude, but this issue and potential solutions require study.

Second, in the preamble of the SP survey questions we have explored (see Figure 2), we ask respondents to imagine that a few aspects of well-being change while all others are held constant. A potential problem is what we call *irrepressible imputation*: when one aspect is varied, respondents may impute variation in a related aspect, in spite of explicit instructions not to do so. For example, when asked to imagine that "your sense that your life is meaningful and has value" increases, respondents might think that "how happy you feel" also increases. Such imputation might occur because the respondent believes that one causes the other or because they are highly correlated in everyday experience. If such imputation occurs, then we, the

econometricians, will obtain a biased estimate of the aspect's MU.

Finally, if policy-makers desire to assess both objective and subjective dimensions of well-being (as in Stiglitz, Sen and Fitoussi, 2009), how can *objective measures* of aspects of well-being best be incorporated into a PWB index? Objective measures are attractive because they eliminate the need for eliciting the aspect's level on the SWB survey. But they introduce a new problem for the SP survey: to correctly evaluate tradeoffs involving an objective aspect, respondents must be able to make accurate judgments about how the objective units (e.g., $\mu\text{g}/\text{m}^3$ of $\text{PM}_{2.5}$ air pollution) relate to their well-being.

To conclude, while we share the enthusiasm of many in government and academia for national well-being measurement, and while we think there is a promising roadmap, we agree with the conclusion of recent reports such as Stone and Mackie (2013) that many obstacles remain. Finding ways to overcome them seems to us an exciting and important research agenda.

REFERENCES

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 Nichole Szembrot. 2014. "Beyond Happiness and Satisfaction: Toward Well-Being Indices Based on Stated Preference." *American Economic Review*, 104(9), 2698–2735.

Deaton, Angus, and Arthur Stone. 2016. "Understanding context effects for a measure of life evaluation: how responses matter." *Oxford Economic Papers*, 68 (4), 861–870.

Fleurbaey, Marc, and Didier Blanchet. 2013. *Beyond GDP: Measuring Welfare and Assessing Sustainability*. Oxford University Press.

Lancaster, Kelvin J. 1966. "A New Approach to Consumer Theory" *Journal of Political Economy*, 74 (2), 132–157.

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-toussi.fr.

Stone, Arthur A., and Christopher Mackie, eds. 2014. *Subjective Well-Being: Measuring Happiness, Suffering, and Other*

Dimensions of Experience. National
Academies Press.