Global Research Unit

Global Research Unit Working Paper #2019-029

Self-reported wellbeing indicators are a valuable complement to traditional economic indicators but aren't yet ready to compete with them

Dan Benjamin, University of Southern California/NBER
Kristen Cooper, Gordon College
Ori Heffetz, Cornell University/NBER
Miles Kimball, University of Colorado Boulder/NBER

CityU 香港城市大學 City University of Hong Kong 專業 創新 胸懷全球 Professional·Creative For The World
商 城大商學院 College of Business CITY UNIVERSITY OF HONG KONG
Department of ECONOMICS & FINANCE

Self-reported wellbeing indicators are a valuable complement to traditional economic indicators but aren't yet ready to compete with them

A Comment on
A happy choice: wellbeing as the goal of government
By Paul Frijters, Andrew E. Clark, Christian Krekel and Richard Layard

Dan Benjamin, Kristen Cooper, Ori Heffetz, and Miles Kimball*

Abstract

We join the call for governments to routinely collect survey-based measures of self-reported wellbeing and for researchers to study them. We list a number of challenges that have to be overcome in order for these measures to eventually achieve a status competitive with traditional economic indicators. We discuss in more detail one of the challenges, *comprehensiveness*: single-question wellbeing measures do not seem to fully capture what people care about. We briefly review the existing evidence, suggesting that survey respondents, when asked to make real or hypothetical tradeoffs, would not always choose to maximize their predicted response to single-question wellbeing measures. The deviations appear systematic, and persist under conditions where alternative explanations are less plausible. We also review an approach for combining single-question measures into a more comprehensive wellbeing index—an approach that itself is not free of ongoing theoretical and implementational challenges, but that we view as a promising direction.

Introduction

We commend Frijters, Clark, Krekel, and Layard for bringing the goal of measuring and raising people's actual wellbeing to the forefront of the policy debate. We salute the work they and others have done, and are continuing to do, in drawing the attention of researchers and policymakers to survey-based wellbeing indicators. We join them in their call for "a flood of experimentation by individuals, firms, communities, and departments in all areas of life to see what improves wellbeing." And we agree with their proposed next steps: governments should systematically collect wellbeing indicators; where possible, policies should be implemented in ways that make causal inference possible; methodologies for using wellbeing indicators for

* Benjamin: Center for Economic and Social Research and Department of Economics, University of Southern California, and NBER (email: djbenjam@usc.edu). Cooper: Department of Economics and Business, Gordon College (email: kristen.cooper@gordon.edu). Heffetz (corresponding author): Department of Economics and Center for Rationality, The Hebrew University of Jerusalem, Samuel Curtis Johnson Graduate School of Management, Cornell University, and NBER (email: oh33@cornell.edu). Kimball: Department of Economics, University of Colorado Boulder, and NBER (email: miles.kimball@colorado.edu). We are grateful for NIH/NIA grant R01-AG051903 to the University of Southern California for financial support; to Matthew Adler, Angus Deaton, and Andrew Oswald for insightful comments on an earlier draft; and to participants in the Third Annual Stockholm Region Workshop on Economics & Philosophy for valuable feedback.

policy analysis should be developed and improved; and government analysts should be trained in the interpretation and use of wellbeing indicators. Over the past decades, much has been learned through research on wellbeing indicators, and we hope to see this research continue to expand.

Where we may disagree is the extent to which current measures of wellbeing are ready for "prime time." While they do not say so explicitly, Frijters et al.'s forceful presentation of the "case for wellbeing" appears to suggest that existing wellbeing indicators—particularly survey data on life satisfaction—should be used *now* to guide policy decisions. Frijters et al. do acknowledge some key limitations of existing knowledge, e.g., "wellbeing regressions yield relatively small income coefficients…the direct implication is a relatively large willingness-to-pay that does not pass the smell test, casting doubt on the use of wellbeing data more generally" (section 7.2). Yet they give the overall impression that the limitations are not impediments to the immediate adoption of wellbeing indicators and their preferential use in policymaking.

In our view, there are many substantial challenges to be overcome before survey-based wellbeing indicators will be ready to stand on par with traditional economic and social indicators. We focus our Comment on one of these challenges that our own research has emphasized: the failure of single-question wellbeing indicators to comprehensively capture wellbeing. As discussed further below, by "wellbeing" we mean idealized choice: what people would choose for themselves if they were fully informed, clear-headed, unbiased, and the like. At the end of our Comment, we briefly outline some of the other major challenges on which we believe substantial progress should also be made before wellbeing indicators are widely adopted for guiding policy.

The non-comprehensiveness of single-question wellbeing indicators

As Frijters et al. eloquently articulate, the core motivation for using survey-based wellbeing indicators is that they hold out the promise of capturing *all* dimensions of subjective experience that people care about—and therefore, according to some thinkers, *everything* that matters for policy. Traditional economic indicators measure only some of what people care about (for example, GDP aims to capture consumption of market goods, investment, and government expenditures), and therefore they can at best serve as non-comprehensive proxies for overall wellbeing. To the extent that a policy has important effects on unaccounted-for dimensions of wellbeing, such as health, relationships, emotions, and other quality-of-life dimensions, reliance on traditional indicators alone can yield bad policy prescriptions.

Do single-question wellbeing indicators, such as questions about happiness or life satisfaction as advocated by Frijters et al., fulfil the promise of being a comprehensive measure of wellbeing? Before we turn to research that addresses this question directly, we note that some of the work that has studied wellbeing indicators may raise doubts (that go beyond the small income coefficients mentioned by Frijters et al.). For example, Stevenson and Wolfers (2009) report a "paradox": despite improvement in their lives on many objective measures, women's happiness in the U.S. declined between 1972 and 2006. There are plausible explanations for this decline—for example, women may be held to higher expectations that are more difficult to

meet—and understanding the reasons for the decline may yield important insights for policy. Nonetheless, for those of us with strong priors that women are better off today than half a century ago, the paradox suggests skepticism about treating happiness self-reports as comprehensive measures of wellbeing.

Evidence from several papers more directly addresses the question of whether wellbeing indicators are comprehensive measures of wellbeing, including two of ours that we briefly describe below (see also Adler and Dolan, 2008; Adler, Dolan, and Kavetsos, 2017; Fleurbaey and Schwandt, 2016; Szabó and Ujhelyi, 2017). Like Adler (2011) and others, our benchmark for assessing whether a measure comprehensively captures wellbeing is *idealized choice*. This benchmark is widely, albeit not universally, accepted in modern moral philosophy. One classic definition is: "an individual's good consists in what he would want himself to want, or to pursue, were he to contemplate his present situation from a standpoint fully and vividly informed about himself and his circumstances, and entirely free of cognitive error or lapses of instrumental rationality" (Railton, 1986, p. 16). Since idealized choice is unobservable, in our empirical work we attempt to find situations, hypothetical and real, where observed choices and stated choices would come as close as possible to such idealized choices.[1]

Our first paper (Benjamin, Heffetz, Kimball, and Rees-Jones, 2012) is nicely summarized by Frijters et al. We face survey respondents with hypothetical choices between two options—for example, a job with higher income vs. a job allowing for more hours of sleep—and ask them two questions: Which of the two options do you think would make you happier (or more satisfied with your life, etc.)? Which of the two do you think you would choose? If people were seeking to maximize their happiness or life satisfaction, then, to the extent that stated choice is a reasonable approximation to idealized choice in the context of the scenarios, our respondents would answer the two questions similarly. We find that they often—but far from always—do, and we identify additional factors—such as predicted sense of purpose, control over one's life, family happiness, and social status—that help explain stated choice.

In our second paper (Benjamin, Heffetz, Kimball, and Rees-Jones, 2014) we replace the hypothetical choices from our first paper with real choice data, collected by surveying 561 students from 23 U.S. medical schools shortly after they submitted their choice rankings over residencies to the National Resident Matching Program. We elicit a list of predictions regarding happiness, life satisfaction, and residency features (such as prestige and future career prospects) for respondents' top-ranking programs, and compare reported program-choice rankings with happiness- or life-satisfaction-prediction rankings. Our evidence from this context, which is real-world and high-stakes and in which the medical students deliberate about their choices, is consistent with our hypothetical-choice evidence from the earlier paper.

[1] The use of choice as a benchmark, under circumstances that are *close* to (but fall short of) idealized choice, is a venerable tradition in economics and has justifications rooted in moral philosophy (e.g., Bernheim, 2016). In contrast, we avoid using revealed choice as a benchmark under circumstances where we suspect it may be far from idealized choice (in the extreme: addiction, suicide). To be clear, idealized choice is conceptually distinct from both of Kahneman, Wakker, and Sarin's (1997) notions of "decision utility" (which describes what people choose, possibly erroneously due to biases or lack of deliberation or information) and "experienced utility" (which describes how people answer questions about current pleasure and pain).

In this second paper, we also provide evidence on the wellbeing-indicator-based approach to cost-effectiveness analysis, advocated by Frijters et al. in their section 6. With our data, we can directly estimate the medical students' willingness to trade off between perceived features of the residency programs by regressing residents' preference rankings—as revealed by the actual residency choices they report to us—on their beliefs about these features, and calculating the ratio of estimated coefficients. We treat these ratios as choice-based estimates of the students' "marginal rates of substitution." We then compare them with estimates obtained by a wellbeing-indicator-based approach: we use the exact same method as above, but instead of regressing actual choice, we regress predicted happiness (or life satisfaction) for a residency program on beliefs about program features, and calculate the ratio of estimated coefficients. We find that these wellbeing-indicator-based ratios are typically different from the choice-based marginal rates of substitution by 60% or more. This evidence—that the coefficients from wellbeing regressions may be rather different from those from (the informed and deliberated) choice regressions—provides concrete evidence in support of Frijters et al.'s concern that the coefficients from wellbeing regressions do not always pass the smell test. We therefore caution against a blanket recommendation to use them in cost-effectiveness analysis.

The findings from these two papers suggest that people care about more than just what is measured by standard, single-question survey measures of "happiness" or even "life satisfaction." Thus, these measures do not solve the basic problem that motivates their use as an alternative to traditional economic indicators. Using such measures to replace more traditional methods could amount to merely replacing one non-comprehensive proxy for wellbeing with another.

One might conjecture that since responses to happiness or life satisfaction questions plausibly capture *more* of what people care about (despite not being fully comprehensive), they must at least *improve on* traditional economic indicators. Whether this conjecture is true depends on whether people's responses to these survey questions accurately weight the dimensions of wellbeing relative to each other. This remains an open question, but some of the evidence is negative. For example, in our hypothetical-choice paper mentioned above (Benjamin, Heffetz, Kimball, and Rees-Jones, 2012), we find that people are more likely to choose options involving higher income than to think that these options will increase their life satisfaction or happiness. This finding suggests that the survey measures underweight the opportunities afforded by a higher income relative to other dimensions of wellbeing—again providing evidence that supports Frijters et al.'s smell-test concern. In other words, while traditional indicators likely overvalue income and consumption, single-question wellbeing indicators may swing too far in the opposite direction, undervaluing them.

An approach to more comprehensive survey-based wellbeing indicators

While effort is ongoing to find a single survey question that would more comprehensively capture wellbeing, another idea—which to us seems more promising at present—is to ask a *set* of survey questions that (to a reasonable approximation) are: (i) more comprehensive when taken all together, and (ii) non-overlapping (or, if overlap is unavoidable,

with the weights used to aggregate the questions adjusted to correct for "double-counting").[2] The set of questions may include single-question measures of emotions (e.g., happiness, anxiety) and broad evaluations (e.g., life satisfaction, Cantril ladder) but also questions about other dimensions of wellbeing that people care about, such as health, sense of purpose, or supporting their families financially. Responses to the set of questions are then aggregated into an overall *wellbeing index* that can be used as a comprehensive survey-based measure of wellbeing. To inform that aggregation, other survey questions are used to elicit relative local preferences over changes in the different dimensions of wellbeing.

We are aware of two approaches to wellbeing indexes grounded in economic theory that have been explored in the literature to date. For concreteness, here we briefly describe one of them (the other is described in, e.g., Fleurbaey and Blanchet, 2013; Decancq, Fleurbaey, and Schokkaert, 2015). The approach we sketch here is detailed in Benjamin, Heffetz, Kimball, and Szembrot (2014) and is summarized and updated in Benjamin, Cooper, Heffetz, and Kimball (2017), from which we borrow much of the text below. It aims to derive an individual-level wellbeing index that can be tracked over time or used as an outcome in policy evaluation. To identify dimensions of wellbeing to measure, we began by constructing a list that includes over 100 potential survey questions, drawing on proposals from psychologists, philosophers, and economists, from Maslow (1946) to Stiglitz, Sen, and Fitoussi (2009) and beyond. We then expanded this list to over 2000 potential dimensions, and proposed ways to eventually shorten to a manageable list that retains what the data tell us are, empirically, the most important dimensions.

The theory for measuring wellbeing is analogous to that underlying the measurement of aggregate consumption. In the case of aggregate consumption, a utility function $u(\boldsymbol{c})$ is defined over a consumption vector for M goods. A traditional aggregate consumption index, $\sum_{m=1}^{M} \overline{p_m}\, c_m$, weights each good's consumption, c_m, by its price held fixed at a baseline level, $\overline{p_m}$. In the face of small changes in consumption, changes in the index approximate changes in utility (up to a multiplicative constant): $\sum_{m=1}^{M} \overline{p_m}\, \Delta c_m \propto \sum_{m=1}^{M} \frac{\partial u(\boldsymbol{c})}{\partial c_m} \Delta c_m \approx \Delta u$.

In the case of a survey-based wellbeing index, the vector of M consumption goods, $\boldsymbol{c}$, is replaced with a vector of J dimensions of wellbeing, $\boldsymbol{w}$. These dimensions are the final goods that people ultimately care about (market goods are now treated as intermediate goods in the production of dimensions of wellbeing, à la Becker, 1965). The levels of $\boldsymbol{w}$ are assumed to be what is measured by responses to the survey questions.

Because the dimensions are not traded in markets, price data are unavailable, and different individuals may have different marginal rates of substitution across the dimensions. Nevertheless, a wellbeing index can be constructed using each individual's marginal rates of substitution for the dimensions as weights. Specifically, the index is given by $\sum_{j=1}^{J} \frac{\overline{\partial u(\boldsymbol{w})}}{\partial w_j} w_j$,

[2] Note that this motivation for asking multiple questions—in order to obtain a more comprehensive measure—is different from (and complementary with) the usual motivation for constructing psychological batteries, which is to minimize measurement error.

where the marginal utilities are defined relative to an arbitrary numeraire dimension. Small changes in this index provide a first-order approximation to changes in the individual's (ordinal) utility (even if the individual's preferences are non-linear).

In practical terms, then, this approach requires: (a) identifying a set of dimensions of wellbeing to measure, (b) asking survey questions that measure each dimension, and (c) asking additional, stated-preference survey questions that measure how respondents would be willing to make trade-offs between the dimensions. The marginal rates of substitution estimated from responses to these stated-preference questions are then used as weights to construct an index that aggregates across the dimensions of wellbeing.

Of course, such an approach to constructing a wellbeing index is data hungry, both in absolute terms and, especially, compared with a single wellbeing question. We certainly acknowledge the *practical* challenges of implementing the index approaches—the higher costs of data-collection requirements that they place on governments, the effort they impose on respondents given the survey length and format, the difficulty of explaining the wellbeing measurement methodology to constituents, etc.—all of which may make it tempting to take the single-question data we already have at present and run with it. However, these practical challenges should not be confused with arguments about what is the *appropriate* methodology for measuring wellbeing in a way that is normatively appealing and theoretically sound.

Other issues for ongoing research

We have highlighted non-comprehensiveness because we think it is a major reason to be cautious about basing policy on single-question wellbeing measures—and because the search for comprehensiveness is an important reason for economists to look "beyond GDP" in the first place. However, there are also a number of other not-yet-resolved challenges to the use of survey-based wellbeing for policy, regardless of whether these are single-question measures or indexes (see also Adler, 2013). To give some examples:

Which question(s) to ask. In their section 5, Frijters et al. raise but do not answer the question of *which* wellbeing indicator should be used for policy purposes. It is widely agreed that different wellbeing measures capture different aspects of experiences, e.g., a happiness question aims to measure emotions, whereas life satisfaction gets at an overall evaluative assessment. These measures have been found to correlate differently with policy-relevant variables (e.g., Kahneman and Deaton, 2010; Stone, Schwartz, Broderick, and Deaton, 2010). For a wellbeing index, it is not yet clear what *set* of questions is sufficiently comprehensive and non-overlapping to be used.

Question-interpretation differences. A related issue is that different people may *interpret* the question being asked differently (Ross, Eyman, and Kishchuk, 1986; Ralph, Palmer, and Olney, 2011; Benjamin, Debnam, Fleurbaey, Heffetz, and Kimball, 2019). For example, some people might interpret a happiness question as asking about their *own* happiness, while others might interpret it as asking about both their own and their family's happiness. If so, different responses to a happiness question might reflect differences in interpretation rather than differences in

wellbeing. Thus, even with standard questions about happiness or life satisfaction, it is worth investigating whether a better-worded question might lead to a more homogeneous interpretation of the question.

Scale-use differences. When asked to answer a wellbeing question on some response scale, say 0-10, different people may use the response scale differently (see Oswald, 2008, on the curvature of the "reporting function from objective reality to subjective feelings"). For example, some sociodemographic groups are less likely to stick to the middle response option than others (see Van Vaerenbergh and Thomas, 2013, for a recent review of this "response style" literature). Thus, when some people report an increase in life satisfaction from 5 to 6, they might actually be experiencing a bigger improvement than other people who report the same increase. One approach to account for differences in scale use uses "vignette" survey questions to try to measure and correct for such differences (e.g., Kapteyn, Smith, and van Soest, 2007), but that approach has been criticized (e.g., Deaton, 2011). Differences in scale-use that are correlated with group differences or caused by events clearly pose a challenge for use of wellbeing indicators subject to these scale-use differences in policy evaluation.

Adaptation. A related phenomenon is adaptation: in response to a positive life event (such as winning the lottery) or negative life event (such as becoming disabled), people's happiness or life satisfaction changes a lot initially but subsequently moves partway back toward its initial level (for a review, see, e.g., Lucas, 2007). Such adaptation may to some extent reflect how feelings of happiness or life satisfaction actually evolve over time, but it also seems to reflect a change in how people *answer questions* about their happiness or life satisfaction (Loewenstein and Ubel, 2008).

Bounded scale and top-coding. The response scale for a wellbeing survey question should allow for wellbeing improvement over time. The bounded scale and the tendency for respondents to use numbers near the top of the scale in standard questions may preclude individuals from reporting genuine improvements.

Other measurement and survey-methodology issues. Due mainly to concerns about a variety of measurement issues, the U.S. National Academy of Sciences "Panel on Measuring Subjective Well-Being in a Policy-Relevant Framework" charged with studying hedonic measures (e.g., happiness questions) concluded that these measures were not yet ready to be treated as national statistics (Stone and Mackie, 2013). There are also survey-methodology concerns. For example, some widely used self-reported wellbeing data are collected through surveys with low response rates (around or below 50 percent). In these cases, conclusions regarding cross-group differences may depend on who self-selects into responding, and on how hard the survey administrators attempt to get nonrespondents to participate (Heffetz and Rabin, 2013). Until the main measurement and survey-methodology issues are addressed, we worry about relying too heavily on estimates from the self-reported-wellbeing literature for making policy decisions.

Aggregating wellbeing indicators across individuals. As Frijters et al. acknowledge and discuss in their section 6.1, the equal utilitarian weights that they assume throughout much of their paper are not normatively compelling. The weights used to aggregate individuals' welfare

to a social welfare function should reflect society's normative views about inequality. Put more precisely, aggregation should reflect a normative consensus about curvature of the social welfare function. Is a half-point increase to someone beginning at 9 on a 0-10 scale really worth the same as a half-point increase to someone beginning at 3? (Top-coding naturally imposes some curvature, but this might not be the right curvature.) For a review of recent developments in social welfare theory, see Fleurbaey and Blanchet (2013).

Conclusion

We share Frijters et al.'s view that survey-based wellbeing indicators could improve policy-making, and improve people's lives. We therefore fully agree with their next steps, including regularly estimating and reporting the effects of policy on wellbeing indicators. We view such analysis of wellbeing indicators as complementary to analysis of standard economic indicators, owing to the likely non-comprehensiveness of each.

Our disagreement is a matter of emphasis. Frijters et al. seem to argue that we should be using wellbeing indicators *now* for policy, and we will figure out how to overcome the challenges as we go. In our view, research on the use of survey-based wellbeing indicators for policy is in its infancy. While we agree that "measurement does not need to be perfect to be useful," we worry that fetishizing an imperfect measure could be damaging—as demonstrated by the obsession with GDP. We would like to see more progress on overcoming some of the measurement and methodological challenges before we are comfortable advocating the use of these less well-understood indicators as a primary means of policy evaluation. We hope that researchers reading Frijters et al.'s contribution will, like us, appreciate it as a call for making that progress in the required additional research. And we hope that policymakers reading their contribution will appreciate it as a call for working closely with researchers on such research.

Understood in this way, we fully join Frijters et al.'s call. Directing national statistics offices to develop and field new wellbeing measures; making sure policy interventions and reforms are designed to help estimate their wellbeing effects; collaborating with researchers on field experimentation—in short, creating the data needed for such research and making it available to the research community—is our vision of the important role policymakers could currently play in bringing about the change that Frijters et al. and we envision.

References

Adler, M.D., 2011. *Well-Being and Fair Distribution: Beyond Cost-Benefit Analysis*. Oxford University Press.

Adler, M. D., 2013. Happiness Surveys and Public Policy: What's the Use? *Duke Law Journal*, 62(8), pp.1509-1601.

Adler, M. D. and Dolan, P., 2008. Introducing a 'Different Lives' Approach to the Valuation of Health and Well-Being. *University of Pennsylvania Institute for Law and Economics, Research Paper Series,* Paper No. 08-05.

Adler, M. D., Dolan, P., and Kavetsos, G., 2017. Would you choose to be happy? Tradeoffs between happiness and the other dimensions of life in a large population survey. *Journal of Economic Behavior and Organization*, 139, pp. 60-73.

Becker, Gary S., 1965. A Theory of the Allocation of Time. *Economic Journal*, 75(299), pp.493-517.

Benjamin, D. J., Cooper, K. B., Heffetz, O., and Kimball, M., 2017. Challenges in Constructing A Survey-Based Well-Being Index. *American Economic Review,* 107(5), pp.81-85.

Benjamin, D. J., Debnam, J., Fleurbaey, F., Heffetz, O., Kimball, M. S., 2019. What Do Happiness Data Mean? Evidence from a Survey of Happiness Respondents. Unpublished manuscript.

Benjamin, D. J., Heffetz, O., Kimball, M. S., and Rees-Jones, A., 2012. What Do You Think Would Make You Happier? What Do You Think You Would Choose? *American Economic Review,* 102(5), pp.2083-2110.

Benjamin, D. J., Heffetz, O., Kimball, M. S., and Rees-Jones, A., 2014. Can Marginal Rates of Substitution Be Inferred from Happiness Data? Evidence from Residency Choices. *American Economic Review,* 104(11), pp.3498-3528.

Benjamin, D. J., Heffetz, O., Kimball, M. S., and Szembrot, N., 2014. Beyond Happiness and Satisfaction: Toward Well-Being Indices Based on Stated Preference. *American Economic Review*, 104(9), pp. 2698-2735.

Bernheim, B. D., 2016. The Good, the Bad, and the Ugly: A Unified Approach to Behavioral Welfare Economics. *Journal of Benefit-Cost Analysis,* 7(1), pp.12-68.

Deaton, A., 2011. Comment on "Work Disability, Work, and Justification Bias in Europe and the US". In *Explorations in the Economics of Aging* (pp. 312-314). University of Chicago Press.

Decancq, K., Fleurbaey, F., Schokkaert, E., 2015. Happiness, equivalent incomes, and respect for individual preferences. *Economica*, 82, 1082–1106.

Fleurbaey, M. and Blanchet, D., 2013. *Beyond GDP: Measuring welfare and assessing sustainability*. Oxford University Press.

Fleurbaey, M. and Schwandt, H., 2016. How some people can maximize their happiness even though they are not actively pursuing it. *USApp– American Politics and Policy Blog,* [blog] 22 June. Available at: <https://blogs.lse.ac.uk/usappblog/> [Accessed 29 June 2019].

Heffetz, O. and Rabin, M., 2013. Conclusions Regarding Cross-Group Differences in Happiness Depend on Difficulty of Reaching Respondents. *American Economic Review,* 103(7), pp. 3001-3021.

Holmstrom, B., and Milgrom, P., 1991. Multitask Principle—Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. *Journal of Law, Economics, and Organization,* 7, pp. 24-52.

Kahneman, D., and Deaton, A., 2010. High income improves evaluation of life but not emotional well-being. *Proceedings of the National Academy of Sciences*, 107(38) pp. 16489-16493.

Kapteyn, A., Smith, J. P., and van Soest, A., 2007. Vignettes and Self-Reports of Work Disability in the United States and the Netherlands. *American Economic Review*, 97(1), pp.461-473.

Loewenstein, G. and Ubel, P. A., 2008. Hedonic adaptation and the role of decision and experience utility in public policy. *Journal of Public Economics*, 92(8-9), pp.1795-1810.

Lucas, R. E. 2007. Adaptation and the set-point model of subjective well-being: Does happiness change after major life events? *Current Directions in Psychological Science*, 16(2), 75-79.

Maslow, A., 1946. A theory of human motivation. In: P. L. Harriman, ed. 1946. *Twentieth Century Psychology: Recent Developments in Psychology.* New York: The Philosophical Library, Inc. pp.22-48.

Oswald, A. J., 2008. On the Curvature of the Reporting Function from Objective Reality to Subjective Feelings. *Economics Letters*, 100(3), pp. 369-372.

Railton, P. 1986. Facts and Values. *Philosophical Topics*, 14(2), pp.5-31.

Ralph, K., Palmer, K., and Olney, J., 2011. Subjective Well-being: a qualitative investigation of subjective well-being questions. Working paper for the Technical Advisory Group on 29 March 2012, Office of National Statistics, U.K.

Ross, M., Eyman, A., and Kishchuk, N., 1986. Determinants of Subjective Well-Being. In James M. Olson, C. P. Herman, and Mark P. Zanna (eds.), *Relative Deprivation and Social Comparison: The Ontario Symposium, Volume 4*. L. Erlbaum Associates.

Stevenson, B., and Wolfers, J., 2009. The Paradox of Declining Female Happiness. *American Economic Journal: Economic Policy,* 1(2), pp. 190-225.

Stiglitz, J. E., Sen, A. and Fitoussi, J. P., 2009. Report by the Commission on the Measurement of Economic and Social Progress. *See* http://www.stiglitz-sen-fitoussi.fr/en/index.htm.

Stone, A., Schwartz, J., Broderick, J., and Deaton, A., 2010. A snapshot of the age distribution of psychological well-being in the United States. *Proceedings of the National Academy of Sciences,* 107(22), pp. 9985-9990.

Stone, A.A., and Mackie, C. (Eds.), 2013. *Subjective Well-Being: Measuring Happiness, Suffering, and Other Dimensions of Experience*. Washington, D.C., U.S.: National Academies Press.

Szabó, A. and Ujhelyi, G., 2017. Choice and Happiness in South Africa. *Economics Letters,* 155, pp.28-30.

Van Vaerenbergh, Y., and Thomas, T. D., 2013. Response Styles in Survey Research: A Literature Review of Antecedents, Consequences, and Remedies. *International Journal of Public Opinion Research*, 25(2), pp.195-217.