Articles

HAPPINESS SURVEYS AND PUBLIC POLICY: WHAT'S THE USE?

MATTHEW D. ADLER†

ABSTRACT

Subjective well-being (SWB) surveys ask respondents to quantify their overall or momentary happiness or life-satisfaction, or pose similar questions about other aspects of respondents' mental states. A large empirical literature in economics and psychology has grown up around such surveys. Increasingly, too, scholars have advanced the normative proposal that SWB surveys be used for policymaking—for example, by using survey results to calculate monetary equivalents for nonmarket goods (to be incorporated in cost-benefit analysis), or to calculate "gross national happiness."

This Article skeptically evaluates the policy role of SWB data. It is critical to distinguish between (1) using SWB surveys as evidence of preference utility versus (2) using them as evidence of experience utility. Preference utility is a measure of the extent to which someone has realized her preferences; experience utility, a measure of the quality of someone's mental states. The two are quite different because individuals can have preferences regarding non-mental occurrences.

Having drawn this distinction, the Article then argues, first, that SWB surveys are poor evidence of preference utility—given problems

of preference and scale heterogeneity, as well as other difficulties. Stated-preference surveys are a much better survey format for eliciting preference utility. Second, in considering SWB surveys as an experience-utility measure, we should recognize that "experientialism" about well-being—the view that well-being is simply a matter of good experiences—is highly controversial. More plausibly, an experience-utility measure might be seen as an indicator of one aspect of well-being. However, even constructing this "weak" experience-utility measure is not straightforward—as the Article demonstrates by discussing Daniel Kahneman's detailed proposal for such a metric.

TABLE OF CONTENTS

INTRODUCTION

"Happiness" is all the rage. A fast-growing literature in economics, growing out of work by research psychologists concerning positive psychological states, examines the determinants of individuals' happiness or feelings of satisfaction with their lives.[1] Much work in this literature, often referred to as the literature on subjective well-being (SWB), is based upon large-scale surveys, posing questions such as, "Taken all together, would you say that you are very happy, pretty happy, or not too happy?" Or, "All things considered, how satisfied would you say you are with your life these days? Please tell me on a scale of 1 to 10, where 1 means very dissatisfied and 10 means very satisfied." A respondent's answer to such a question can then be correlated with information about other attributes, ascertained by additional questions—for example, her income, relative income, employment status, whether she is married or has children, her health condition, and so forth.

---

1. For overviews, see generally LUIGINO BRUNI & PIER LUIGI PORTA, ECONOMICS AND HAPPINESS: FRAMING THE ANALYSIS (2005); BRUNO S. FREY, HAPPINESS: A REVOLUTION IN ECONOMICS (2008); BRUNO S. FREY & ALOIS STUTZER, HAPPINESS AND ECONOMICS: HOW THE ECONOMY AND INSTITUTIONS AFFECT WELL-BEING (2002); HANDBOOK ON THE ECONOMICS OF HAPPINESS (Luigino Bruni & Pier Luigi Porta eds., 2007); THE SCIENCE OF WELL-BEING (Felicia A. Huppert, Nick Baylis & Barry Keverne eds., 2005); WELL-BEING: THE FOUNDATIONS OF HEDONIC PSYCHOLOGY (Daniel Kahneman, Ed Diener & Norbert Schwarz eds., 1999); Ed Diener, Eunkook M. Suh, Richard E. Lucas & Heidi L. Smith, *Subjective Well-Being: Three Decades of Progress*, 125 PSYCHOL. BULL. 276 (1999); Paul Dolan, Tessa Peasgood & Mathew White, *Do We Really Know What Makes Us Happy? A Review of the Economic Literature on the Factors Associated with Subjective Well-Being*, 29 J. ECON. PSYCHOL. 94 (2008).

DUKE LAW JOURNAL [Vol. 62:1509

Table 1 excerpts the questions about happiness or life satisfaction that are posed by the surveys most widely used by SWB researchers.[2]

Table 1. Widely Used SWB Questions

Survey	Variable	Question
General Social Survey	Happiness	Taken all together, how would you say things are these days? Would you say that you are very happy, pretty happy, or not too happy?
World Values Survey	Life sat.	All things considered, how satisfied are you with your life as a whole these days? Please use this card to help with your answer. [range of 1-10 with 1 labelled "Very Dissatisfied" and 10 labelled "Very Satisfied"]
European Social Survey	Happiness	Taking all things together, how happy would you say you are? Please use this card [range of 0-10 with 0 labelled "Extremely unhappy" and 10 labelled "Extremely happy"]
European Social Survey	Life sat.	All things considered, how satisfied are you with your life as a whole nowadays? Please answer using this card, where 0 means extremely dissatisfied and 10 means extremely satisfied. [range of 0-10 with 0 labelled "Extremely dissatisfied" and 10 labelled "Extremely satisfied"]
European Quality of Life Survey	Happiness	Taking all things together on a scale of 1 of 10, how happy would you say you are? Here 1 means you are very unhappy and 10 means you are very happy.
European Quality of Life Survey	Life sat.	All things considered, how satisfied would you say you are with your life these days? Please tell me on a scale of 1 to 10, where 1 means very dissatisfied and 10 means very satisfied.
German Socio-Economic Panel	Life sat.	In conclusion, we would like to ask you about your satisfaction with your life in general. Please answer according to the following scale: 0 means 'completely dissatisfied', 10 means 'completely satisfied'. How satisfied are you with your life, all things considered?
British Household Panel Survey	Life sat.	How dissatisfied or satisfied are you with your life overall? [range of 1-7 with 1 labelled "Not satisfied at all" and 7 labelled "Completely satisfied".]

Some SWB researchers have objected to these standard questions, which invite respondents to express their *overall* happiness or life satisfaction, that is, to reflect on all the circumstances of their lives, and then to quantify their happiness or life satisfaction with these total circumstances in view. Daniel Kahneman, most prominently, has argued that SWB surveys should focus instead on moment-to-moment happiness.[3] The trajectory of an individual's momentary happiness over a given time period can be estimated via "experience sampling"[4]: she will be given a small electronic device which will periodically beep, prompting her to quantify her happiness at that moment. Alternatively, and somewhat less intrusively, a survey might employ the so-called "day reconstruction method."[5] At the end of each day during the survey period, respondents will be asked to recollect the episodes of the day and to rate the affective quality of each episode.[6]

Kahneman and collaborators used the day-reconstruction method in a study of working women in Texas.[7] Respondents were asked to assign each episode a number from 0 (not at all) to 6 (very much) on various scales—for example, how "happy" she felt during the episode, how "warm/friendly," how "frustrated/annoyed," or how "depressed/blue."[8] A measure of the respondent's "positive affect" during the episode was calculated by averaging how she rated the episode on three of the scales (happy, warm/friendly, enjoyment), and her "negative affect" during the episode was similarly calculated by averaging her ratings on six other scales (frustrated/annoyed, depressed/blue, hassled/pushed around, angry/hostile, worried/anxious, criticized/put down).

The questions posed by Kahneman and collaborators in the Texas study are certainly different, in some respects, from the questions summarized in Table 1. For that matter, "How happy are

---

3. *See infra* note 176; *see also infra* Part III.B.

4. Alan B. Krueger, Daniel Kahneman, David Schkade, Norbert Schwarz & Arthur A. Stone, *National Time Accounting: The Currency of Life*, *in* MEASURING THE SUBJECTIVE WELL-BEING OF NATIONS: NATIONAL ACCOUNTS OF TIME USE AND WELL-BEING 9, 30–31 (Alan B. Krueger ed., 2009).

5. *Id.*

6. *Id.*

7. Daniel Kahneman, Alan B. Krueger, David A. Schkade, Norbert Schwarz & Arthur A. Stone, *A Survey Method for Characterizing Daily Life Experience: The Day Reconstruction Method*, 306 SCIENCE 1776, 1776–77 (2004).

8. *Id.* at 1777 tbl.1.

you?" is not semantically equivalent to "How satisfied are you with your life?" We might expect to see (and indeed *do* see) a somewhat different pattern of responses to happiness versus life-satisfaction questions.[9] Still, questions regarding momentary happiness, momentary affects, overall happiness, or overall life satisfaction share a crucial similarity: all ask the respondent to consider some aspect of her mental life and to rate that aspect on a numerical scale or, nearly equivalently, to assign that aspect to one of a series of ordered categories. For short, I will call any such survey a "SWB" survey and data or research grounded on such surveys "SWB" data or research. Note that my definition includes the surveys understood by researchers as paradigmatic SWB surveys, namely the surveys in Table 1, as well as questions such as, "How angry are you right now?," "How painful is this?," or "How happy did you feel when you were doing that?"

Scholars working in the SWB literature tend to be empiricists— psychologists or empirical economists who use sophisticated econometric techniques to tease out the causal connections that SWB data might illuminate. However, by no means has the literature been wholly empirical and non-normative. Strikingly, many of the leading researchers have also advanced *normative* recommendations regarding government policy.[10] They have drawn upon SWB research

---

9. *See* Ed Diener, Weiting Ng, James Harter & Raksha Arora, *Wealth and Happiness Across the World: Material Prosperity Predicts Life Evaluation, Whereas Psychosocial Prosperity Predicts Positive Feeling*, 99 J. PERSONALITY & SOC. PSYCHOL. 52, 58–60 (2010); Daniel Kahneman & Angus Deaton, *High Income Improves Evaluation of Life but Not Emotional Well-Being*, 107 PROC. NAT'L ACAD. SCI. 16,489, 16,489 (2010).

10. Scholarship endorsing the use of SWB surveys for law or policy purposes includes DEREK BOK, THE POLITICS OF HAPPINESS: WHAT GOVERNMENT CAN LEARN FROM THE NEW RESEARCH ON WELL-BEING (2010); ED DIENER, RICHARD E. LUCAS, ULRICH SCHIMMACK & JOHN F. HELLIWELL, WELL-BEING FOR PUBLIC POLICY (2009); FREY, *supra* note 1; FREY & STUTZER, *supra* note 1; CAROL GRAHAM, THE PURSUIT OF HAPPINESS: AN ECONOMY OF WELL-BEING (2011); RICHARD LAYARD, HAPPINESS: LESSONS FROM A NEW SCIENCE (rev. & updated ed. 2011); MEASURING THE SUBJECTIVE WELL-BEING OF NATIONS, *supra* note 4; Samuel R. Bagenstos & Margo Schlanger, *Hedonic Damages, Hedonic Adaptation, and Disability*, 60 VAND. L. REV. 745 (2007); John Bronsteen, Christopher Buccafusco & Jonathan S. Masur, *Welfare as Happiness*, 98 GEO. L.J. 1583 (2010) [hereinafter Bronsteen, Buccafusco and Masur, *Welfare as Happiness*]; John Bronsteen, Christopher Buccafusco & Jonathan S. Masur, *Well-Being Analysis vs. Cost-Benefit Analysis*, 62 DUKE L.J. 1603 (2013) [hereinafter Bronsteen, Buccafusco and Masur, *Well-Being Analysis*]; Ed Diener & Martin E.P. Seligman, *Beyond Money: Toward an Economy of Well-Being*, 5 PSYCHOL. SCI. PUB. INT. 1 (2004); Ed Diener, *Subjective Well-Being: The Science of Happiness and a Proposal for a National Index*, 55 AM. PSYCHOLOGIST 34 (2000); Paul Dolan & Mathew P. White, *How Can Measures of Subjective Well-Being Be Used To Inform Public Policy?*, 2 PERSP. PSYCHOL. SCI. 71 (2007); Peter Henry Huang, *Happiness Studies and Legal Policy*, 6 ANN. REV. L. & SOC. SCI. 405

either to advocate particular policy measures or to argue for new kinds of policy-evaluation tools.

Two such tools have received sustained scholarly attention. The first is the use of SWB data to determine *monetary equivalents* for nonmarket goods, with these equivalents then incorporated in cost-benefit analysis.[11] Imagine that regression techniques are used to estimate an equation from survey data, whereby an individual's life satisfaction measured on a numerical scale is the dependent variable and various independent variables are given, including *both* an individual's income *and* her level of some nonmarket good. The coefficient on income is c_y, meaning that an increase in income by \$1 is estimated to produce c_y units of increase in life satisfaction, whereas the coefficient on the nonmarket good is c_G, meaning that an increase of 1 unit in an individual's level of the good is estimated to produce c_G units of increase in life satisfaction. The monetary equivalent for 1 unit of the good is thus c_G/c_y dollars.[12]

(2010); and Daniel Kahneman, Alan B. Krueger, David Schkade, Norbert Schwarz & Arthur Stone, *Toward National Well-Being Accounts*, 94 AM. ECON. REV. (PAPERS & PROC.) 429 (2004). For further such scholarship, see *infra* notes 11, 13.

    11.   For overviews of this approach, see DANIEL FUJIWARA & ROSS CAMPBELL, DEP'T FOR WORK & PENSIONS (UK), VALUATION TECHNIQUES FOR SOCIAL COST-BENEFIT ANALYSIS: STATED PREFERENCE, REVEALED PREFERENCE AND SUBJECTIVE WELL-BEING APPROACHES (2011); Andrew E. Clark & Andrew J. Oswald, *A Simple Statistical Method for Measuring How Life Events Affect Happiness*, 31 INT'L J. EPIDEMIOLOGY 1139 (2002); Bruno S. Frey, Simon Luechinger & Alois Stutzer, *The Life Satisfaction Approach to Environmental Valuation*, 2 ANN. REV. RESOURCE ECON. 139 (2010); Heinz Welsch & Jan Kühling, *Using Happiness Data for Environmental Valuation: Issues and Applications*, 23 J. ECON. SURVS. 385 (2009); and Paul Dolan, Daniel Fujiwara & Robert Metcalfe, *A Step Towards Valuing Utility the Marginal and Cardinal Way* (Ctr. for Econ. Performance, Discussion Paper No. 1062, 2011). Recent examples include Christopher L. Ambrey & Christopher M. Fleming, *Valuing Scenic Amenity Using Life Satisfaction Data*, 72 ECOLOGICAL ECON. 106 (2011); Arik Levinson, *Valuing Public Goods Using Happiness Data: The Case of Air Quality*, 96 J. PUB. ECON. 869 (2012); Tobias Menz & Heinz Welsch, *Life-Cycle and Cohort Effects in the Valuation of Air Quality: Evidence from Subjective Well-Being Data*, 88 LAND ECON. 300 (2012); and Nattavudh Powdthavee & Bernard van den Berg, *Putting Different Price Tags on the Same Health Condition: Re-Evaluating the Well-Being Valuation Approach*, 30 J. HEALTH ECON. 1032 (2011).

    12.   An increase in income by c_G/c_y dollars produces $c_y(c_G/c_y) = c_G$ increase in happiness, and thus is the equivalent, in happiness terms, of an increase in exposure to the good by 1 unit.

         Bronsteen, Buccafusco and Masur have recently proposed that policies be evaluated via "well-being analysis" (WBA), whereby policy impacts, including income changes, would be converted into happiness units. *See* Bronsteen, Buccafusco and Masur, *Welfare as Happiness*, *supra* note 10, at 1627–40; Bronsteen, Buccafusco and Masur, *Well-Being Analysis*, *supra* note 10, at Part II. Although this proposal has some similarities to the use of SWB data to calculate monetary equivalents for nonmarket goods, it differs from the latter in important respects. First, Bronsteen, Buccafusco, and Masur adopt the "experience-quality" (EQ) defense of SWB surveys, indeed the strong EQ defense—to use a distinction I will develop later in the Article,

Substantial scholarly work along these lines has been undertaken, using SWB data to estimate monetary equivalents for various goods and bads—including air quality, airport noise, other environmental goods, the death of family members, social relationships, exposure to the risk of crime or terrorism, and unemployment.

Second, scholars have worked toward developing an SWB-survey-based index of social condition akin to gross domestic product (GDP).[13] (The desire to develop such an index is, in part, what has motivated Kahneman's research.) The idea here is to have a measure of individual happiness that can be added across persons.[14] We could then estimate the gross national happiness or average happiness of a country at a given time. We could also look at time trends in gross/average happiness, characterize international or inter-group differences in gross/average happiness, and use such a measure to assess particular policies by predicting the change in gross/average happiness that these policies would yield.

Governments have also begun to gather SWB data and to initiate serious bureaucratic and even political discussions about the possibility of SWB-survey-based policymaking. The Kingdom of Bhutan has long endorsed the concept of "Gross National

---

see infra Part I.C. They see WBA as a way to take account of the information that SWB surveys provide about individuals' experience utility. By contrast, scholars using the monetary-equivalent approach do *not* universally adopt the EQ defense, at least not explicitly so. One *possible* reading of some of this scholarship is that it adopts the "preference-realization" (PR) defense: that the monetary-equivalent approach is meant to be sensitive to the information SWB surveys provide concerning *preference utility*. *See infra* Part I.C. Indeed, Part II discusses the monetary-equivalent approach, at length, as a test case for the PR defense of SWB surveys.

Second, WBA requires cardinal happiness data; the monetary-equivalent approach requires only ordinal data. *See* Bronsteen, Buccafusco and Masur, *Well-Being Analysis*, *supra* note 10, at Part II.B.4; *infra* Part II.A.3. Finally, WBA neutralizes wealth effects, whereas (in principle) the monetary-equivalent approach does not. To see this point, imagine that some nonmarket good has the very same effect on SWB for both rich and poor, and that money has a diminishing incremental effect. Then WBA will be neutral between a policy that provides the good for free to some number of rich individuals and one that does so for the same number of poor individuals, taxing some third population, whereas the monetary-equivalent approach will (in principle) favor the first policy.

    13.   *See generally* MEASURING THE SUBJECTIVE WELL-BEING OF NATIONS, *supra* note 4; Diener & Seligman, *supra* note 10; Kahneman et al., *supra* note 10; Ruut Veenhoven, *Happy Life-Expectancy: A Comprehensive Measure of Quality-of-Life in Nations*, 39 SOC. INDICATORS RES. 1 (1996); Ruut Veenhoven & Wim Kalmijn, *Inequality-Adjusted Happiness in Nations: Egalitarianism and Utilitarianism Married in a New Index of Societal Performance*, 6 J. HAPPINESS STUD. 421 (2005).

    14.   For discussion of cardinality and ordinality, see *infra* Part II.A.3.

Happiness" and has used it in guiding economic development.[15] Until recently, such happiness-oriented governance seemed highly idiosyncratic. In 2008, however, the president of France, Nicholas Sarkozy, commissioned the influential Stiglitz-Sen-Fitoussi report to investigate non-GDP measures of social welfare. The report recommended (inter alia) SWB surveys:

> Research has shown that it is possible to collect meaningful and reliable data on subjective as well as objective well-being. Subjective well-being encompasses different aspects (cognitive evaluations of one's life, happiness, satisfaction, positive emotions such as joy and pride, and negative emotions such as pain and worry): each of them should be measured separately to derive a more comprehensive appreciation of people's lives. Quantitative measures of these subjective aspects hold the promise of delivering not just a good measure of quality of life per se, but also a better understanding of its determinants, reaching beyond people's income and material conditions. Despite the persistence of many unresolved issues, these subjective measures provide important information about quality of life. Because of this, the types of question that have proved their value within small-scale and unofficial surveys should be included in larger-scale surveys undertaken by official statistical offices.[16]

The Cameron government in the United Kingdom has welcomed the Stiglitz-Sen-Fitoussi report,[17] and statistical bureaus in the United Kingdom are now implementing a program of SWB surveys.[18] This also appears to be in the offing in France and in the United States.

But do SWB surveys really offer a new foundation for governmental policy? In this Article, I critically interrogate the

---

15.   *See* Winton Bates, *Gross National Happiness*, 23 ASIAN-PAC. ECON. LITERATURE 1, 12 (2009).

16.   *See* JOSEPH E. STIGLITZ, AMARTYA SEN & JEAN-PAUL FITOUSSI, REPORT BY THE COMMISSION ON THE MEASUREMENT OF ECONOMIC PERFORMANCE AND SOCIAL PROGRESS 16 (2009), *available at* http://www.stiglitz-sen-fitoussi.fr/documents/rapport_anglais.pdf.

17.   PAUL DOLAN, RICHARD LAYARD & ROBERT METCALFE, OFFICE FOR NAT'L STAT., MEASURING SUBJECTIVE WELL-BEING FOR PUBLIC POLICY 3 (2011).

18.   *See* PAUL DOLAN & ROBERT METCALFE, OFFICE FOR NAT'L STAT., COMPARING MEASURES OF SUBJECTIVE WELL-BEING AND VIEWS ABOUT THE ROLE THEY SHOULD PLAY IN POLICY (2011); DOLAN ET AL., *supra* note 17 at 3; OFFICE FOR NAT'L STAT., FIRST ANNUAL ONS EXPERIMENTAL SUBJECTIVE WELL-BEING RESULTS (2012), *available at* http://www.ons.gov.uk/ons/dcp171766_272294.pdf; SAM WALDRON, OFFICE FOR NAT'L STAT., MEASURING SUBJECTIVE WELLBEING IN THE UK (2010).

normative recommendations of SWB researchers (and the Stiglitz-Sen-Fitoussi report).[19]

A key difficulty is a lack of rigor and care in handling relevant normative concepts, such as "well-being," "preference satisfaction," and "utility." Relatedly, the SWB literature has largely failed to engage with contemporary philosophical scholarship on well-being. To be sure, this lack of engagement is reciprocal, and it is not particularly surprising, given academic specialization. SWB research has been spearheaded, as mentioned, by psychologists and empirically minded economists. The standard methods of moral philosophers—conceptual analysis, intuition-pumping via simple and unrealistic thought experiments—are unfamiliar in these social-scientific realms.

But now that SWB scholarship has moved beyond the boundaries of social science, from the explanatory to the normative—to a posture of *endorsing* particular governmental policies, policy-evaluation approaches, or SWB measures as the basis for such approaches—its methods require supplementation too. A more precise and nuanced engagement with different possible conceptions of well-being is needed. This Article tries to exemplify, and in any event seeks to encourage, such engagement.

Part I reviews the philosophical literature, seeking to clarify crucial distinctions that are all too often blurred by SWB scholars: between well-being and a particular, experientialist, conception of well-being (for example, the view that well-being is just happiness); between the satisfaction of someone's preferences (how well actual conditions in the world fit with her preference ranking) and the fact that the individual *feels* satisfied; between utility in the traditional economic sense, a measure of preference satisfaction, and experience utility, a measure of the individual's happiness, positive affects, or feelings of satisfaction.

---

19. For other, important, critical work on SWB surveys, see the chapters by George Loewenstein, William Nordhaus, and Erik Hurst in MEASURING THE SUBJECTIVE WELL-BEING OF NATIONS, *supra* note 4; MARC FLEURBAEY & DIDIER BLANCHET, BEYOND GDP: MEASURING WELFARE AND ASSESSING SUSTAINABILITY (forthcoming 2013) (manuscript at ch. 5) (on file with the *Duke Law Journal*); Daniel M. Hausman, *Hedonism and Welfare Economics*, 26 ECON. & PHIL. 321 (2010); Mark Kelman, *Hedonic Psychology and the Ambiguities of "Welfare,"* 33 PHIL. & PUB. AFF. 391 (2005); George Loewenstein & Peter A. Ubel, *Hedonic Adaptation and the Role of Decision and Experience Utility in Public Policy*, 92 J. PUB. ECON. 1795 (2008); Rick Swedloff & Peter H. Huang, *Tort Damages and the New Science of Happiness*, 85 IND. L.J. 553 (2010); and Marc Fleurbaey, Erik Schockkaert & Koen Decancq, *What Good Is Happiness?* (Ctr. for Operations Res. & Econometrics, Discussion Paper No. 17, 2009).

Part I concludes by differentiating two quite distinct roles that SWB surveys might play as inputs to policy choice. First, a SWB survey might evidence the degree to which an individual's preferences are satisfied. I will term this the "preference-realization" (PR) defense of SWB surveys. Let it be understood (a point I will repeat below) that, on a standard preference-based account of well-being, the *fundamental arguments* for an individual's preferences—what she intrinsically prefers and disprefers—might well include items external to her mind, such as her physical health, her accumulation of material goods, her freedom and autonomy, her accomplishments, and so forth. Second, an SWB survey might indicate the quality of an individual's mental states. It might tell us whether her mental states are good or bad in some sense, for example, whether she feels happy or distressed. I call a normative defense of SWB surveys predicated on *this* role the "experience-quality" (EQ) defense.

Kahneman, clearly and decisively, presses the EQ defense. He sees a (properly constructed) SWB survey as indicating whether the individual is currently experiencing a positive or negative affective state and, in either event, how intense it is. Other researchers, however, seem to see SWB surveys as evidence of preference realization. Still others offer both defenses, often without acknowledging that the two involve quite different accounts of *what* information SWB surveys are supposed to be providing.

The distinction between PR and EQ defenses of the policy relevance of SWB data clarifies normative discussion. It is a distinction *with* a difference—a distinction that SWB scholars would do well to notice—and a distinction that will serve as the organizing template for this Article. In Part II, I consider, and reject, the PR defense. In Part III, I consider the EQ defense—here drawing a further line between "strong" and "weak" variants. The strong variant of the EQ defense, adopted by Richard Layard in his influential book on happiness,[20] adopts an experientialist view of well-being. Experientialism about well-being is a controversial view. It says that, necessarily, an individual is equally well-off in two outcomes if she has the same mental states in both (even if the condition of her physical body, or facts about the wider world, are different in the two outcomes).

---

20.   LAYARD, *supra* note 10. For a description and critique of Layard's views, see *infra* Part III.A.

The weak variant of the EQ defense, now favored by Kahneman, acknowledges the possibility of nonexperiential constituents of well-being, but stresses that happiness or, more generally, mental states, are at least *one* important aspect of human flourishing.[21] The weak EQ defense is by far the most plausible basis for incorporating SWB data into policy analysis. Even *this* defense of the policy relevance of SWB surveys, however, is vulnerable to substantial criticisms. I will flesh out these criticisms via a close analysis of Kahneman's work. Part III will conclude by articulating a *different* methodology for taking account of the impact of governmental policies on individuals' mental states, a methodology that does not rely upon SWB surveys.

The aims of the Article, thus, are both conceptual and substantive. First, it articulates distinctions that are useful in thinking carefully about the normative relevance of SWB surveys—above all, the distinction between PR and EQ defenses. Second, it argues substantively that neither the PR defense nor the EQ defense, in even its weak form, is convincing. The second aspect of the Article is separable from the first. Even if the reader is not persuaded by my various critiques of SWB surveys, she will still agree (I hope) that it is vital to differentiate the potential role of SWB surveys as evidence of *preference* utility (the PR defense) from their role as evidence of *experience* utility (the EQ defense).[22]

---

21. *See infra* notes 177–178 and accompanying text.

22. I am grateful to Carol Graham for writing a commentary on this Article. Carol Graham, *An Economist's Perspective on Well-Being Analysis and Cost-Benefit Analysis*, 62 DUKE L.J. 1691 (2013). Graham distinguishes between two types of SWB surveys: hedonic well-being (HWB) surveys (which seek to measure individuals' affects) and evaluative well-being (EWB) surveys (which seek to measure how individuals evaluate their lives). *Id.* at 1692–93. To be clear, my intention in this Article is not to defend HWB surveys as contrasted with EWB surveys. Part II focuses on life-satisfaction surveys (a kind of EWB survey) as a comparatively more plausible indicator of preference utility than happiness surveys—but my aim in undertaking the analysis of Part II is to demonstrate that the PR defense with respect to SWB surveys *of any kind* is problematic. (Because EWB surveys fail in that role, as Part II shows, then a fortiori HWB surveys do.) Part III, in discussing the weak EQ defense, focuses on Kahneman's "objective happiness" framework. It does so because Kahneman's work is the most fully developed version of the weak EQ defense. As it happens, "objective happiness" relies on HWB surveys (asking about momentary affects) rather than EWB surveys. I do not mean to suggest that the weak EQ approach is *best* fleshed out by focusing wholly on affects, and indeed I criticize Kahneman for doing just that. *See infra* Part III.B.5.

I. PREFERENCES, EXPERIENCES, AND THE NORMATIVE ROLE OF SUBJECTIVE WELL-BEING SURVEYS

Section A of this Part clarifies some important well-being concepts. A key distinction is that between *experientialist* and *nonexperientialist* conceptions of well-being. A preference-based conception of well-being falls in the nonexperientialist category because an individual can hold an intrinsic preference for items other than her own experiences. For example, she might prefer to be in good health, to have children, or to attain various goals she has set for herself, *as such*, rather than merely preferring the experiences that flow from health, children, or goal fulfillment.

Section B looks at the empirical evidence regarding individuals' intrinsic preferences for nonexperiential items. If such preferences were empirically rare, policy-analysis tools might safely ignore them. But the available evidence does not demonstrate that this is the case.

Finally, building upon the conceptual tools set forth here, Section C differentiates between PR (preference-realization) and EQ (experience-quality) defenses of the policy role of SWB surveys.

A. *Well-Being, Happiness, Preference, Experience, and Utility: Some Clarifications*

In an article published in the prestigious *Economic Journal*, Richard Easterlin—who pioneered the use of SWB data in economics—writes, "Throughout this article, I use the terms happiness, subjective well-being, satisfaction, utility, well-being, and welfare interchangeably."[23] For anyone acquainted with the vibrant contemporary philosophical literature on welfare, Easterlin's equivalences—and similar conflations drawn throughout the SWB literature[24]—are unsettling. Imagine how a labor or financial

---

    23.   Richard A. Easterlin, *Income and Happiness: Towards a Unified Theory*, 111 ECON. J. 465, 465 (2001).

    24.   *See, e.g.*, DIENER ET AL., *supra* note 10, at 9–12 (suggesting that "economic theories of well-being equate well-being with utility," that "economists define utility as the satisfaction that a person experiences from the consumption of goods," and that the authors' proposed subjective definition of well-being, in terms of individuals' favorable self-evaluations, is "essentially identical to economists' concept of utility"); FREY, *supra* note 1, at 3 ("In general, . . . as in the literature, the terms 'happiness', 'well-being', and 'life satisfaction' are used interchangeably."); David G. Blanchflower & Andrew J. Oswald, *Well-Being over Time in Britain and the USA*, 88 J. PUB. ECON. 1359, 1360–62 (2004) (using well-being, happiness, and utility as equivalent terms); Diener & Seligman, *supra* note 10, at 1 (defining "[w]ell-being" as "peoples' positive evaluations of their lives"); Andrew J. Oswald, *Happiness and Economic Performance*, 107 ECON. J. 1815, 1815 (1997) (leaping from the premise that economic

economist would react to the statement, "I use the terms income, consumption, and wealth interchangeably."

Academic philosophers writing about well-being standardly draw a series of distinctions that I will now articulate. Are these distinctions "right" or "true" or "correct" in some foundational sense? Maybe not. For purely predictive or explanatory purposes, it might be perfectly appropriate to equate "happiness, subjective well-being, satisfaction, utility, well-being, and welfare," as Easterlin does.[25] And asserting the "correctness," for normative purposes, of a scheme that disaggregates well-being, happiness, preference satisfaction, and utility—a disaggregation commonplace in contemporary philosophical writing about welfare[26]—implicates contested metaethical questions that I am quite happy (!) to avoid here.

I will observe, however, that academic philosophers—the scholarly community engaged in serious, longstanding, normative debates about well-being—have found it useful to deploy a certain vocabulary. This has served to make precise plausible normative views. And it seems pretty risky for SWB scholars, now also engaged in normative debates, to ignore the vocabulary or to dumb it down.

First, philosophers typically recognize that there is a *plurality* of accounts of well-being. One such account sees an individual's well-being as based upon various objective goods. Another equates well-being and preference attainment. A third account, the hedonic account, dating back to Jeremy Bentham, says that an individual's

---

performance has no intrinsic normative significance, to the conclusion that "[e]conomic things matter only in so far as they make people happier"); Carol D. Ryff & Burton H. Singer, *Know Thyself and Become What You Are: A Eudaimonic Approach to Psychological Well-Being*, 9 J. HAPPINESS STUD. 13, 14–15 (2008) (viewing Aristotelian "eudaimonia" as a kind of psychological state, namely "psychological well-being"); Bernard M.S. van Praag, *Perspectives from the Happiness Literature and the Role of New Instruments for Policy Analysis*, 53 CESIFO ECON. STUD. 42, 42 & n.1 (2007) (assuming that "economic behaviour . . . is motivated by maximization of utility, satisfaction, well-being or happiness" and noting that "[w]e will make no difference between these notions"). For further examples of SWB scholars using the concepts of well-being, subjective well-being, and/or happiness as equivalent, see Erik Angner, *Are Subjective Measures of Well-Being 'Direct'?*, 89 AUSTRALASIAN J. PHIL. 115, 119–20 (2011).

   25.   *See supra* note 23 and accompanying text.

   26.   *See, e.g.*, DANIEL M. HAUSMAN, PREFERENCE, VALUE, CHOICE, AND WELFARE 77–81 (2012); DANIEL M. HAYBRON, THE PURSUIT OF UNHAPPINESS: THE ELUSIVE PSYCHOLOGY OF WELL-BEING 29–42 (2008); L.W. SUMNER, WELFARE, HAPPINESS, AND ETHICS 81–98 (1996); Erik Angner, *Subjective Well-Being*, 39 J. SOCIO-ECON. 361, 362–63 (2010).

welfare depends upon her pains and pleasures. And there are further possibilities.[27]

Choosing between these accounts is a matter for substantive normative argument. Purely semantical or definitional considerations will not suffice to vindicate one account. Sheer reflection on the meaning of the word "well-being" will not establish, for example, that a preference-based view of well-being is correct, and that those who espouse competing views are in error. Instead, the proponent of any particular account will need to bring into play the ordinary tools of normative debate: showing that this account has intuitively appealing results for particular cases, that it fits with more general principles that seem attractive, and so forth.

Second, philosophers categorize the plurality of accounts of well-being in various ways (with such categorization seen as illuminating normative debate). One such categorization is binary, demarcating between accounts that make an individual's well-being wholly dependent upon the quality of her mental states, and accounts that do not have this feature. I will make this demarcation precise by distinguishing between *experientialist* and *nonexperientialist* accounts of welfare:

> *Experientialist versus Nonexperientialist Accounts of Well-Being: A Definition*
> An experientialist account of well-being satisfies the following requirement: if a given individual's mental states (her beliefs, affects, memories, desires, and all other aspects of her mental states) in outcome x are identical to her mental states in outcome y, and outcomes x and y are maximally specified,[28] then the

---

27. For an overview of the philosophical literature on well-being, see MATTHEW D. ADLER, WELL-BEING AND FAIR DISTRIBUTION: BEYOND COST-BENEFIT ANALYSIS 158–85 (2012).

28. An "outcome" is a possible state of affairs: a description of some aspect of the world, which might possibly occur or have occurred in the past, present, and/or future. A maximally specified outcome is what philosophers term a "possible world," namely an outcome F such that for every other F^* either F entails F^* or F entails not-F^*. In effect, a maximally specified outcome is a complete history of the universe, leaving no possible occurrence undetermined.

Of course, maximally specified outcomes are not items that humans can hold in consciousness. In particular, human policy analysts cannot use maximally specified outcomes to think about policy impacts. Rather, maximally specified outcomes are theoretical constructs that serve various theoretical purposes. For example, they help to categorize different accounts of well-being. An experientialist account says that the quality of an individual's experiences is the only intrinsic determinant of her well-being; changes in nonexperiential items can *cause* changes in well-being, but do not change it directly. This feature of experientialism is made precise by

individual is equally well-off in both outcomes. A nonexperientialist account is any account that fails to satisfy the experientialism requirement. (A nonexperientialist account says that it is possible for there to be an individual, *i*, and maximally specified outcomes *x* and *y*, such that *i* is better off in *x* than *y* even though all of the individual's mental states are the same in the two outcomes.)

A hedonic account of well-being is experientialist in the sense defined here. Assume that all aspects of a given individual's mental states are identical in (maximally specified) outcomes *x* and *y*. Then, in particular, her pain sensations in *x* are identical to her pain sensations in *y*. And her pleasure sensations in *x* are identical to her pleasure sensations in *y*. If so, someone who adopts a hedonic conception of welfare will say that the individual is equally well-off in the two outcomes.

The hedonic account is *not* the sole member of the class of experientialist views. A richer account, also experientialist, says that an individual's well-being depends not only upon her pains and pleasures, but also upon her attainment of high quality *nonhedonic* mental states (for example, having good memories, cognitions, or perceptions) and her avoidance of low quality *nonhedonic* mental states.[29] It is easy to see why this richer account is still experientialist as per the above definition. Assume that *all* aspects of an individual's mental life are identical in (maximally specified) outcomes *x* and *y*: her pains and pleasures *and* her memories, cognitions, perceptions,

---

the definition in the text only if *x* and *y* are maximally specified. Even an experientialist account might allow the welfare ranking of *partially described* outcomes to be a partial function of nonmental attributes (given their *causal* role).

For example, assume that outcomes are partially described to specify an individual's income and how much pain or pleasure she feels, but not the nonhedonic aspects of her mental life (for example, her sense of life satisfaction or the quality of her memories). Assume that individual *i*'s pains and pleasures are the same in *x* and *y*, but that she has more income in *x*. Then someone who reduces well-being to pains/pleasures *and* a sense of life satisfaction—a kind of sophisticated experientialist account—might say that individual *i* is better off in *x* than *y* despite the fact that all of *i*'s specified mental attributes are identical in *x* and *y*. Having greater income might cause a change in unspecified, well-being-relevant mental attributes, namely, a sense of life satisfaction.

By contrast, if *x* and *y* are fully specified, the causal upshots of any nonmental attribute that an individual might possess are already "built into" the description of the outcomes, and so the only reason for a difference in some nonmental attribute to yield a difference in well-being is because of the attribute's intrinsic welfare significance.

29. *See infra* notes 197–200 and accompanying text.

and so forth. Then, even on the account which says that well-being is a matter of attaining pleasure *and* high-quality nonhedonic states, and avoiding pain *and* low-quality nonhedonic states, the individual is equally well-off in the two outcomes.

The concept of "happiness" is standardly (although perhaps not invariably) specified in experientialist terms.[30] If an individual's affects and cognitions and everything else about her mental life are identical in two outcomes, then it must be the case that she is equally *happy* in both. Happiness (however more precisely defined) is "in the head"—at least on the standard view.

It thus follows that "well-being" and "happiness" are distinct concepts. There is nothing conceptually incoherent in acknowledging that happiness satisfies the experientialism requirement, while simultaneously adopting a nonexperientialist account of welfare, which says that well-being does not satisfy this requirement. Again, nonexperientialism about well-being might be substantively unattractive—it might lose out after normative debate—but such an account is in no way confused or incoherent regarding the very meaning of "well-being."

What, then, *are* important examples of nonexperientialist accounts of welfare? "Objective good" accounts posit a plurality of well-being goods that include but are not limited to good mental states, and that are seen as having some basis other than an individual's preferences. Such accounts are nonexperientialist. Consider, for example, Martha Nussbaum's well-known list of human "capabilities," a modern exemplar of the objective-good approach.[31] The list includes: life; bodily health; bodily integrity; the senses, imagination and thought; emotions; practical reason; affiliation; other species; play; and control over one's environment. Note that a few of these items individually satisfy the experientialism requirement, but that others do not (and thus the overall account does not either). Contrast, for example, emotions and body health.

It is straightforward that an objective-good account of welfare falls within the category of nonexperientialist accounts. A more subtle point—and one absolutely central to this Article—is that preference-

---

30. *See* HAYBRON, *supra* note 26, at 30.

31. MARTHA C. NUSSBAUM, WOMEN AND HUMAN DEVELOPMENT: THE CAPABILITIES APPROACH 78–80 (2000); *see also* Martha C. Nussbaum, *Who Is the Happy Warrior? Philosophy Poses Questions to Psychology*, 37 J. LEGAL STUD. S81 (2008) (critically evaluating proposals to orient policy around SWB).

based accounts are *also* nonexperientialist. A preference-based account ascribes to a given individual i a ranking R_i of outcomes. Depending on the account, R_i might be i's actual preferences, or R_i might be "laundered" in various ways. In either event, what makes R_i a "preference" ranking is (1) that R_i is minimally well-behaved (in particular, it is transitive), and (2) that R_i is connected to individual i's choices. The preference-based view then says that one outcome is better than a second outcome for individual i if and only if the first outcome is ranked higher by R_i.

The "arguments" for an individual's preference ranking are the features of outcomes that move them up or down in the ranking. We can further distinguish between an individual's *intrinsic* preferences and her instrumental preferences. I intrinsically prefer something if I want it for its own sake; I instrumentally prefer something if I want it because I believe it helps me advance my intrinsic preferences. For example, I might have an intrinsic goal to sail around the world, and want to be healthy not because I care about health as such, but because being healthy will enable me to fulfill this goal. I will use the term "fundamental argument" to mean the features of outcomes that an individual intrinsically prefers.

And now we can see why a preference-based account of well-being does not satisfy the experientialism requirement. On such a view, it is quite possible that the fundamental arguments for individual i's preferences *include features of outcomes other than individual i's mental states*. If so, individual i can be better off in one (maximally specified) outcome than in a second even though his mental states are identical in the two outcomes.

Understand that my intention is simply to *explain* what a preference-based view of well-being involves. I am not arguing, here, in favor of such a view, as opposed to a hedonic account or some other account within the experientialist family.[32] My aim, at this

---

32. I have elsewhere endorsed a preference-based view of well-being. *See* MATTHEW D. ADLER & ERIC. A. POSNER, NEW FOUNDATIONS OF COST-BENEFIT ANALYSIS 25–61 (2006); ADLER, *supra* note 27, at 155–231. And the discussion in Part III.C—suggesting a methodology for incorporating information about experiential quality into policy—does presuppose that a preference-based view is more attractive than competing views. However, the remainder of the Article does not commit itself to this view. The nature of well-being is *contested*, and this Article hopes to address a wider readership than persuaded preferentialists. The aim of this Part is clarificatory: to bring to light distinctions between the various well-being accounts that are often obscured in the SWB literature. Part II argues that a potential defense of SWB surveys, in light of a preference-based view (the PR defense), is unpersuasive. This is an argument about the relation between the preference-based view and SWB surveys—not a defense of that view. Part

juncture, is simply to articulate the demarcation set forth by the experientialism requirement and to clarify that preference-based views fall on the nonexperientialist side of that demarcation.

The point can be further clarified by introducing the concept of a utility function. A utility function, in traditional economics, is a real-valued mathematical function $u_i(.)$, which *represents* the preference ranking R_i of individual i.[33] Under standard technical conditions, R_i will be thus representable, so that outcome x is ranked higher by R_i than outcome y if and only if $u_i(x) > u_i(y)$.

If outcomes are maximally specified, the fundamental arguments for an individual's preferences are the entries in her utility function. Let us assume for the moment—simply to facilitate the exposition—that outcomes are maximally specified. We can now see very easily why a preference-based view is not experientialist. Why? It is *possible* that i's utility function takes the form $u_i(M_i, N_i)$ where M_i denotes facts about i's mental states (her pains and pleasures, memories, cognitions, and so forth), and N_i are other facts about individual i or the wider world. If two outcomes are identical with respect to the M_i facts but not with respect to the N_i facts, individual i's utility function may assign a different number to the two outcomes. This will indicate that the outcomes are not ranked equally by R_i; and a preference-based view will then say that the individual is not equally well-off in the outcomes, thus violating the experientialism requirement.

In actual practice, economists and policy analysts do not work with maximally specified outcomes (which are cognitively intractable). The analytical nexus between the fundamental arguments for an individual's preferences (the items she intrinsically prefers or disprefers) and the entries in her utility function becomes more complex. Where outcomes are not maximally specified, a utility function of the form $u_i(M_i, N_i)$ or $u_i(N_i)$—where N_i are nonmental facts—*might* just mean that i has an instrumental preference regarding those facts. The individual prefers various nonmental occurrences because they will cause her to have various experiences. It is also *possible*, however, that i has an intrinsic preference regarding nonmental features of the world and that a utility function of the form $u_i(M_i, N_i)$ or $u_i(N_i)$ is capturing this intrinsic preference.

III.A criticizes experientialism about well-being—but note that "objective good" as well as preference-based views reject experientialism.

33. *See* HAUSMAN, *supra* note 26, at 13–14; ANDREU MAS-COLLEL, MICHAEL D. WHINSTON & JERRY R. GREEN, MICROECONOMIC THEORY 6–9, 46–50 (7th prtg. 2009).

One objection, pressed by readers of early drafts of this Article, is that it does not "make sense" for preferences to have non-mental-state fundamental arguments. In traditional consumer theory, the inputs to an individual's utility function are the physical quantities of various kinds of goods that an individual possesses and consumes (physically interacts with). Now, it might be objected that an *intrinsic* preference to possess and consume various goods embodies an irrational consumption fetishism. Someone might plausibly have an *instrumental* preference to consume bundle *C* rather than *C**, believing that consuming the first bundle will cause her to experience more happiness than consuming the second. But to prefer bundle *C* as such—independent of any belief that having *C* will lead to different experiences than *C**—seems unintelligible.[34]

To think carefully about this objection, we should have in view the different kinds of preference-based accounts, that is, the different ways in which preferences might be "laundered."[35] One kind of view requires simply that preferences satisfy formal rationality conditions (either the minimum requirements constitutive of a preference, or additional requirements). But formal rationality conditions clearly *permit* preferences with non-mental-state fundamental arguments. For example, someone might well have a transitive, complete, continuous, convex, monotonic, and otherwise formally well-behaved ranking of consumption bundles.[36]

A more fully "laundered" account requires that preferences be "intelligible": that they "make sense," rather than merely satisfy formal rationality conditions. Although intrinsic preferences for consumption bundles, or for income, may well fail this requirement, there are various other kinds of intrinsic preferences for nonmental items that seem perfectly intelligible.[37] (1) *Health*. Health economists typically include an individual's health (meaning not merely her

---

34. Amartya Sen famously argues for a closely related proposition, namely that social assessment should focus on the achieved "functionings" that are caused by income together with individuals' physical and social attributes, or opportunities to function, and not income itself. *E.g.*, AMARTYA SEN, INEQUALITY REEXAMINED 28–30 (1992); *see also* Amartya Sen, *Capability and Well-Being*, *in* THE QUALITY OF LIFE 30, 41 (Martha Nussbaum & Amartya Sen eds., 1993) ("Since income is not desired for its own sake, any income-based notion of poverty must refer—directly or indirectly—to those basic ends which are produced by income as means.").

35. *See* ADLER, *supra* note 27, at 170–72.

36. *See* MAS-COLLEL ET AL., *supra* note 33, at 41–50.

37. *See* Dan Moller, *Wealth, Disability, and Happiness*, 39 PHIL. & PUB. AFF. 177, 186–89 (2011).

mental health, but also the physical integrity and functioning of her body) as one of the entries in her utility function.[38] Although this functional structure could merely be capturing an instrumental preference for health, an intrinsic desire to be healthy also seems intelligible—no less so than an intrinsic preference for certain affects, cognitions, or memories.[39] Jim desires not to feel pain; Jane desires not to have a limb amputated (independent of whatever pain limb loss causes). Where is the argument that Jim's desires make sense but Jane's desires do not? (2) *Liberty*. Having more liberties may reduce the quality of someone's experiences.[40] For example, more liberties might cause more frustration at unrealized expectations, or more time spent agonizing over choices.[41] Still, an individual might view liberty as an aspect of personhood; she might intrinsically prefer to increase her liberties "because that's what autonomy means," even recognizing the uncertain or negative experiential impact of this increase. (3) *Goal fulfillment*. Individuals, as they grow to adulthood, develop a wide range of life goals: to have a particular kind of career; to develop mastery of a particular sport, musical instrument, or art form; to travel to certain places; to learn some body of knowledge; to make an impact on the community in a certain way; and so forth. Surely someone can intelligibly defend her preference for fulfilling some particular life goal by saying, "This is just what I've always wanted to do, and have worked for years to accomplish." And, if that were not enough, she could add, "What is it to be an autonomous person, but to develop and pursue life goals?" It would be astonishing

---

38. *See* Béatrice Rey & Jean-Charles Rochet, *Health and Wealth: How Do They Affect Individual Preferences?*, 29 GENEVA PAPERS ON RISK & INS. THEORY 43, 44–46 (2004).

39. For a review of evidence that individuals have an intrinsic preference for health, see *infra* notes 76–84 and accompanying text; *see also* Carol Graham, *Happiness and Health: Lessons—and Questions—for Public Policy*, 27 HEALTH AFF. 72, 73–74 (2008) (reviewing evidence showing imperfect correlation between health and happiness); Carol Graham, Lucas Higuera & Eduardo Lora, *Which Health Conditions Cause the Most Unhappiness?*, 20 HEALTH ECON. 1431, 1432–33 (2011) (same); Peter A. Ubel & George Loewenstein, *Pain and Suffering Awards: They Shouldn't Be (Just) About Pain and Suffering*, 37 J. LEGAL STUD. S195 (2008) (arguing for the intelligibility of an intrinsic preference not to be physically injured).

40. *See* Betsey Stevenson & Justin Wolfers, *The Paradox of Declining Female Happiness*, 1 AM. ECON. J.: ECON. POL'Y 190, 191 (2009) (finding declining female happiness during a period of increased labor market opportunities and other options for women).

41. *See* Carol Graham & Stefano Pettinato, *Frustrated Achievers: Winners, Losers and Subjective Well-Being in New Market Economies*, J. DEV. STUD., Apr. 2002, at 100, 117–20 (providing evidence of "frustrated achievers": upward income mobility shifts income expectations and reduces happiness). *See generally* BARRY SCHWARTZ, THE PARADOX OF CHOICE: WHY MORE IS LESS (2004).

DUKE LAW JOURNAL   [Vol. 62:1509

if the only sensible explanation for why someone might want to master cello, to become an airplane pilot, or to climb Mount Everest were because of what those accomplishments would make her feel, think, or remember. (4) *Knowledge.* Personhood involves both "practical rationality" (the capacity to form and act upon goals) and "theoretical rationality" (the capacity to acquire knowledge).[42] Aristotle thought that the exercise of theoretical rationality was the highest form of human well-being.[43] Whether or not this is true, it is certainly intelligible to have an intrinsic preference for knowledge, a preference to believe true propositions. But note that the degree to which someone possesses knowledge is *not* a wholly experiential fact about her. It depends both upon what she believes, and whether what she believes is *true*, that is, whether it corresponds to facts in the world.[44] (5) *Relationships.* Just as someone might intelligibly see her status as a *human being* as the basis for an intrinsic preference for physical health; and her status as an *autonomous person* as the basis for an intrinsic preference to possess various liberties, to develop and fulfill various life-goals, or to possess knowledge; so, too, someone might see her status as a *social animal* as justifying intrinsic preferences for having friendships, for being a parent, or for having other kinds of relationships. In fact, the existing evidence from SWB surveys does not establish that parenthood increases SWB.[45] If someone believes that being a parent will, on balance, produce some *decrease* in his happiness and feelings of satisfaction, but still prefers to be a parent, is his preference necessarily unintelligible?

Robert Nozick's famous discussion of the "experience machine" is relevant, here:

---

42.   THOMAS HURKA, PERFECTIONISM 38–41 (1993); *see also* JOHN FINNIS, NATURAL LAW AND NATURAL RIGHTS 59–80 (1980) (arguing for the intrinsic value of knowledge); George Loewenstein, *That Which Makes Life Worthwhile*, *in* MEASURING THE SUBJECTIVE WELL-BEING OF NATIONS, *supra* note 4, at 87, 96–97 (same).

43.   Richard Kraut, *Aristotle's Ethics*, STANFORD ENCYCLOPEDIA OF PHILOSOPHY (Mar. 29, 2010), http://plato.stanford.edu/entries/aristotle-ethics.

44.   Moreover, knowledge may not produce happiness or life satisfaction. For evidence that SWB need not increase with education, see Dolan et al., *supra* note 1, at 99–100; and Joop Hartog & Hessel Oosterbeek, *Health, Wealth and Happiness: Why Pursue a Higher Education?*, 17 ECON. EDUC. REV. 245, 251–54 (1998).

45.   *See* Dolan et al., *supra* note 1, at 107 (finding mixed evidence of SWB impact of having children); Thomas Hansen, *Parenthood and Happiness: A Review of Folk Theories Versus Empirical Evidence*, 108 SOC. INDICATORS RES. 29, 44 (2012) (finding that resident children tend to reduce SWB).

> Suppose there were an experience machine that would give you any experience you desired. Superduper neuropsychologists could stimulate your brain so that you would think and feel you were writing a great novel, or making a friend, or reading an interesting book. All the time you would be floating in a tank, with electrodes attached to your brain. Should you plug into this machine for life, preprogramming your life's experiences? If you are worried about missing out on desirable experiences, we can suppose that business enterprises have researched thoroughly the lives of many others. You can pick and choose from their large library or smorgasbord of such experiences . . . . Would you plug in?[46]

My intrinsic preference for some nonmental state of affairs is a preference that this state of affairs actually occur and *not* merely a preference that I believe (perhaps falsely) in such occurrence and have the feelings caused by this belief (perhaps false). If Tim intrinsically prefers to be in good health, then what Tim wants is actually to be free of diseases, not merely to believe that he is disease-free. If Sarah intrinsically prefers to have a loving and faithful spouse, then what Sarah prefers is a spouse who actually loves her, not merely one who succeeds in persuading her that he does (but cheats behind her back). If Emily wants to be an airline pilot, then what she prefers is a complex state of affairs including her own physical movements (that she actually move through the air by piloting an aircraft) and social recognition (that she be recognized by others as a pilot), and not merely that she believe all this. What Nozick's "experience machine" underscores is that some such preferences are perfectly intelligible.

More specifically, what seems intelligible is a preference structure with both nonmental and mental fundamental arguments. In wanting a good marriage, what I want is (1) a certain kind of relationship (including my spouse actually loving me, being faithful, and so forth, as opposed to my being deceived on this score), and (2) being aware of, and made happy by, this relationship.[47] A preference just to have a certain kind of relationship—wholly independent of how happy or miserable it makes me feel—does seem quite odd. In general, it *does* seem quite odd for someone to be wholly indifferent

---

46.   ROBERT NOZICK, ANARCHY, STATE, AND UTOPIA 42–43 (1974).

47.   *Cf.* SUMNER, *supra* note 26, at 139 (analyzing well-being as "authentic happiness," where authenticity requires inter alia that the individual be factually well-informed); Shelly Kagan, *Well-Being as Enjoying the Good*, 23 PHIL. PERSP. 253, 255 (2009) (suggesting that well-being may consist of taking pleasure in the attainment of objective goods).

to her own happiness, pains and pleasures, memories, cognitions, or other mental states. But that is not the issue here. Rather, I am trying to show how a preference-based view of well-being violates the experientialism requirement. It is intelligible to have preferences with *hybrid* fundamental arguments: for some person to care intrinsically that (1) certain nonmental facts occur, and (2) that she have certain beliefs, memories, cognitions, or feelings as a result of those non-mental facts.

The ranking of outcomes in light of such hybrid preferences does not satisfy the experientialism requirement because the preferences have nonmental as well as mental fundamental arguments. And an intelligibility constraint on preferences—at a minimum—will surely allow some such hybrid preferences. It follows that a preference-based view of well-being—*even if* it incorporates an intelligibility constraint—is nonexperientialist.

Yet a different kind of laundering requires fully informed preferences. But there is no reason to think that fully informed preferences must satisfy the experientialism requirement any more than to think that intelligible preferences must.

So much for the objection that properly laundered preferences (preferences that are formally rational, intelligible, and/or fully informed) must satisfy the experientialism requirement. A second objection to my characterization of preference-based accounts of well-being as nonexperientialist might be articulated as follows: "Someone's preferences are relevant to her well-being only if those preferences are *self-interested*. And a self-interest constraint, properly understood, *does* require that preferences satisfy the experientialism requirement." I will postpone discussion of this objection until later in the Article.[48] To anticipate that discussion: although an experientialist conception of "self-interested" has some plausibility, I believe that the preferentialist should reject it. This conception is counterintuitively narrow, and lacks a theoretical warrant, except on a controversial view about personal identity.

In sum, an individual's preferences—even her laundered, self-interested preferences—can have nonmental fundamental arguments. And thus preference-based accounts of well-being, like objective-good accounts—and *unlike* accounts that reduce well-being to pains/pleasures, to happiness, or to richer packages of affects,

---

48.   *See infra* Part III.A.

cognitions, memories, perceptions, and/or other types of mental states—constitute nonexperientialist conceptions of human welfare.

Some readers might remain puzzled by my characterization of preference-based views as nonexperientialist. These readers might think: "A preference is itself a kind of mental state. If individual i's well-being depends upon his preferences, then his well-being depends upon his mental states. So how can a preference-based account of well-being be nonexperientialist?"

Of course, someone's preferences *are* one aspect of her mental states. It would be absurd to say otherwise. However, preference-based views are nonexperientialist *in the specific sense expressed by the experientialism requirement*. That requirement is framed to capture an important feature of well-being accounts, one that is highly relevant to normative debates about them. If an account fails the experientialism requirement, then *it is possible to directly change someone's well-being without changing any of his mental states*. And indeed this is *true* of preference-based accounts. Imagine that individual i has the very same preferences, R_i, in (maximally specified) outcomes x and y; that R_i has nonmental fundamental arguments N_i as well as mental fundamental arguments M_i; that the individual's mental states are identical in the two outcomes (not just his preferences, but his pains and pleasures, cognitions, memories, and everything else about his mental states); but that outcome x is ranked by R_i above outcome y because the N_i facts are different. Then this is a case where i prefers x to y, and is *better off* in x than in y according to a preference-based view, even though everything about his mental states in the two outcomes (including his preferences) are identical.[49]

---

49. A different objection is that any person j whose preferences have non-mental-state fundamental arguments can be mimicked by a "doppelganger" $j*$ who cares only about her mental states. In particular, whenever j intrinsically prefers nonmental fact F, $j*$ has an intrinsic preference concerning her *beliefs*; namely, $j*$ prefers that she believe F. Thus, policy modelers can ignore the possibility of preferences with non-mental-state fundamental arguments—instead "translating" apparent intrinsic preferences for nonexperiential items into intrinsic preferences for the corresponding beliefs.

But one implication of Nozick's "experience machine" example is that j and $j*$ will not necessarily make identical choices. If the machine will cause the false belief F, then $j*$ might enter the machine while j refuses. Moreover, third parties might act very differently toward j and $j*$. If j prefers to have a faithful spouse, and a friend observes that j's spouse is unfaithful, then the friend might tell j but not do the same in $j*$'s case. If j wants to be in good physical health, the government might fund certain health interventions that it would not fund in the case of $j*$, who simply wants to believe she is in good health.

To say that someone's preferences *can* have non-mental-state fundamental arguments is not to say that they *must*. There is nothing in the preference-based account of welfare to *require* that someone intrinsically care about items other than her own mental states. But, conversely, the view does not *prohibit* her from doing so. What someone prefers, and thus what improves her well-being, is for her to determine. On a preference-based account of welfare, it is *possible* for someone—possible, if she has an intrinsic preference for health, liberty, goal fulfillment, knowledge, relationships, or other nonmental features of outcomes—to be directly benefited or harmed without any change in her mental states, by virtue of events that change the N_i facts but not the M_i facts.

Unfortunately, this feature of preference-based views is not always well-understood by researchers in the SWB literature. One tendency in this literature is to adopt an experientialist account of well-being *without substantive normative argument*. This tendency assumes, from the get-go, that well-being is solely a matter of attaining high quality mental states and avoiding low quality ones, whether specified as pain/pleasure, happiness/unhappiness, feelings of satisfaction, or in some other way.[50]

In particular, SWB scholars too often assume that individuals' intrinsic preferences *must* be for mental items.[51] But to do this is to misunderstand the antipaternalism that constitutes an attraction of the preference-based account. For the preferentialist, the extent to which an individual's welfare covaries with her experiences, as opposed to covarying with nonexperiential facts about her (or the wider world), is an *empirical* question—one to be resolved not by scholarly stipulation or assumption, but only by looking to what the individual prefers.

A related confusion concerns the meaning of "utility." As already explained, a preference is a *ranking*, and a utility function a mathematical device for representing that ranking. Again, utility function $u_i(.)$ represents individual i's preferences, R_i, if the following

---

50.   *See supra* note 24 and accompanying text.

51.   *See, e.g.*, FREY, *supra* note 1, at 5 ("[H]appiness is undoubtedly an overriding goal in most people's lives."); LAYARD, *supra* note 10, at 124 (arguing that happiness-maximization "values what people want for themselves, for their children and for their fellow citizens," namely "their happiness"); Daniel J. Benjamin, Ori Heffetz, Miles S. Kimball & Alex Rees-Jones, *What Do You Think Would Make You Happier? What Do You Think You Would Choose?*, 102 AM. ECON. REV. 2083, 2107 (2012) (noting that the implicit view in much of the economics of happiness literature is that SWB is the sole argument for idealized preferences).

is true: whenever outcome x is ranked above outcome y by R_i, $u_i(x) > u_i(y)$.

An individual's utility, in this sense, can change without *any* change in an individual's feelings, beliefs, or perceptions. If $u_i(.)$ includes nonmental states as its entries, taking the form $u_i(M_i, N_i)$ or $u_i(N_i)$, with N_i nonmental features of individual i or the wider world, then individual i can have higher or lower utility in one outcome than in a second outcome even though her beliefs, feelings, enjoyments, and so forth, are exactly the same in both.

In particular, individual i's actual utility—the numerical value that $u_i(.)$ attaches to the actual outcome—can be higher or lower than her utility in some counterfactual outcome without individual i being aware of this. If R_i ranks x over y, and outcome x occurs, it does not follow that individual i realizes that an outcome preferred to y has occurred or realizes that her utility level is higher in the actual outcome than if y had occurred. Think of R_i as an individualized criterion for assessing and comparing different possible (more or less completely specified) histories of the world. In general, i will not be infallible about the extent to which the actual world matches up to this criterion.

This is most clearly true if R_i has nonmental fundamental arguments. Imagine that I prefer to be liked by my peers. In a world y where they laugh behind my back, without my realizing it, I am lower in the R_i ranking than in an otherwise identical world x where I am genuinely liked, and my utility $u_i(y)$ is lower than utility $u_i(x)$; but if y were to occur, I would not perceive that my utility is lower than $u_i(x)$. Indeed, as discussed in the margin, even if the fundamental arguments for R_i are just mental states, individual i will not generally be infallible about his utility level; his beliefs and feelings about his utility need not correspond to his actual utility level.[52]

---

52.   This can occur insofar as individuals incorrectly predict the causal impact of nonmental attributes. Imagine that individual i cares only about his pains and pleasures. If x and y are (incompletely specified) outcomes that describe individuals' income, health, and various other nonhedonic attributes, then it could be the case that $u_i(x) > u_i(y)$, because i will end up with more pleasure in x, but that i incorrectly believes his bundle of nonhedonic attributes in y will cause him more pleasure. (This is just a kind of hedonic forecasting error.) In particular, an individual who cares only about pains and pleasures might give an incorrect answer to a question (such as the following) asking him to compare his preference-utility in the actual world x to what his utility would be in y, given the counterfactual bundle of nonhedonic attributes specified in y. "Consider your actual income and health. Is your preference-utility higher than it would be if, instead, you had this level of health and this much income?"

Economists and philosophers sometimes use the term "satisfaction" with reference to a preference structure.[53] Outcome x better "satisfies" individual i's preferences than y if x is ranked above y by R_i. "Satisfaction," in this sense, does not necessarily have an experiential component: Joe's preferences can be more fully satisfied without Joe *feeling* more satisfied (for just the reasons discussed in the previous two paragraphs). Of course, in ordinary English, the word "satisfy" does have affective connotations, and so instead I will use the term "realization" or "attainment" to mean movement up a preference ranking. Preference-based accounts of well-being reduce the well-being ranking of outcomes, for some person, to the *realization* of that person's preferences—not to the feelings of satisfaction experienced by that person.

The SWB literature regularly blurs the line between preference-realization (on the one hand) and feelings of satisfaction or beliefs regarding preference-realization (on the other), and the related point that individuals can have intrinsic preferences for items other than their own mental states. For example, Ed Diener, Richard Lucas, Ulrich Schimmack, and John Helliwell analyze well-being in terms of individuals' beliefs and self-evaluations: "A life is going well only if the individual who lives this life endorses it as good and evaluates it positively."[54] They then explain:

> Our definition of well-being . . . is clearly a subjective one. People have well-being only when they believe that their life is going well, regardless of whether that life has pleasure, material comforts, a sense of meaning, or any other *objective* feature that has been specified as essential for well-being. . . .
>
> We should note in this context that our subjective definition of well-being is essentially identical to economists' concept of utility. . . . [E]conomists define utility as the satisfaction that a person experiences from the consumption of goods.[55]

Actually, the proposed account of well-being is not "essentially identical to economists' concept of utility." Economists, traditionally, do *not* define "utility" as "the satisfaction that a person experiences from the consumption of goods," but rather as a numerical function representing the extent to which R_i is realized (*whatever* its

---

53. On what it means to "satisfy" preferences, see generally HAUSMAN, *supra* note 26.

54. DIENER ET AL., *supra* note 10, at 11 (citation omitted).

55. *Id.* at 11 (citation omitted).

arguments, be they physical bundles of commodities, health, leisure, or happiness). We increase an individual's utility, in the economic sense, by acting to produce those features of the world that she prefers; utility goes up when this happens. What she experiences or believes as a consequence is a separate question.

One of the great virtues of Kahneman's work on SWB has been his consistent attempts to keep in focus the difference between utility (in the traditional economist's sense) and good experiences. Kahneman highlights this difference by using the terms "decision utility" and "experienced utility." For example, Paul Dolan and Kahneman write:

> The word 'utility' has two distinct meanings: it can refer either to the hedonic experience of an outcome or to the preference or desire for that outcome. These have been labelled experienced utility and decision utility, respectively. Jeremy Bentham first defined utility in hedonic terms, as a measure of pleasure and pain, and economists followed that usage until the twentieth century. . . .
>
> Economists abandoned experienced utility early in the twentieth century, in favour of a new interpretation, in which utility represents 'wantability.' A person's decision utilities are revealed by her choices. . . . Neoclassical welfare economics rests on a concept of decision utility that is cleansed of any reference to hedonic experience . . . .[56]

However, the term "decision utility" as a measure of preference realization is somewhat problematic. It suggests, misleadingly, that a ranking of choices (decisions) is primary, when in fact the standard formal apparatus—expected-utility theory—makes someone's ranking of choices derivative from her ranking of outcomes, plus probabilities. Moreover, although prior generations of economists (impressed by the observability of preferences for choices) *did* favor the use of choice data to infer preferences for outcomes, many economists now happily do so via *surveys* ("stated preference" formats).[57]

Instead, modifying half of Kahneman's terminology, I will distinguish between *preference utility* and *experience utility*. An individual's *preference* utility is a number assigned to outcomes (or

---

56.   Paul Dolan & Daniel Kahneman, *Interpretations of Utility and Their Implications for the Valuation of Health*, 118 ECON. J. 215, 215 (2008) (citations omitted).

57.   For a discussion of both the stated-preference methodology and the traditional view favoring behavioral evidence of preferences, see sources cited *infra* notes 119, 121.

other items, for example, choices) representing her ranking of outcomes (or her ranking of those other items). An individual's *experience* utility is a number assigned to outcomes (or other items[58]) measuring the quality of her experience.

B. *Do Individuals Prefer More Than Good Experiences?*

If preferences with nonmental fundamental arguments were empirically rare, policy analysis might be simplified by ignoring them. Policy models and tools might be structured on the premise that individuals intrinsically care only about their own experiences—with the justification that this simplifies the methodologies without too much loss in accuracy. But there is *not* clear evidence that individuals generally lack an intrinsic preference for health, liberty, goal fulfillment, consumption, relationships, knowledge, or other items that are (at least partly) nonexperiential. Absent such clear evidence, the preferentialist about well-being should endorse policy tools that *leave open the possibility of preferences with nonmental fundamental arguments*.

The relative importance of experiential and nonexperiential arguments in individuals' preference rankings is, in fact, a topic that scholars have *not* systematically investigated. To be sure, the SWB literature has documented, in great detail, the various ways in which individuals fail to maximize their own happiness. However, such findings—without more—do not do much to illuminate the structure of individual preferences. For example, the fact that someone chooses to procreate even though the average parent is no happier than the average childless adult is consistent both with the hypothesis that she has an intrinsic preference for parenting, *and* with the hypothesis that she has an intrinsic preference only for happiness plus mistaken beliefs about the hedonic benefits of parenting.[59] Similarly, someone's decision to pursue a higher education even though education appears to have little hedonic benefit might be motivated by an intrinsic preference for knowledge, or by a misunderstanding about how happy the better educated are.[60]

In order to determine whether an individual intrinsically prefers some nonexperiential feature of outcomes, we need to specify the experiential consequences of different alternatives presented to the

---

58. For example, Kahneman assigns experience utility to moments. *See infra* Part III.B.
59. *See supra* note 45 (describing evidence of SWB impact of parenting).
60. *See supra* note 44 (describing evidence of SWB impact of education).

individual (or estimate her beliefs concerning those upshots), and *then* see what her ranking of the alternatives is. Only a few studies have attempted to do this, and their findings are mixed.

Laura King and Christie Napa tested for the relative strength of individual preferences for happiness and a sense of meaning (types of experiences), as opposed to income, by showing respondents an ostensible "career survey," which respondents believed to be a questionnaire filled out by a subject describing her job.[61] In these questionnaires, the subject characterized her income as relatively high (>$100,000) or low ($21,000 to $30,000), and also answered questions about how happy she was in her job and how meaningful she found the job. Respondents were then asked to rate the desirability of the subject's life—specifically, via three questions, one of which asked, "How much would you like to have this person's life?" on a scale from 1 to 5, with the other two asking about the quality of the subject's life and the extent to which she was living "the good life." King and Napa found that the subject's happiness and sense of meaning were much more significant than her income in driving respondents' desirability ratings.[62] However, the relative importance of happiness and sense of meaning, as opposed to *non*income items, was not examined.

In preliminary, unpublished work, Paul Dolan and I presented respondents with different hypothetical lives—characterized in terms of the subject's income, life expectancy, health, and happiness (specifically, what proportion of the time the subject was in a good mood)—and asked respondents to identify their most preferred life, their second choice, and so forth. We found the subject's health, and then happiness, to be the most significant features in explaining respondents' preferences over the hypothetical lives.[63]

Daniel Benjamin, Ori Heffetz, Miles Kimball, and Alex Rees-Jones studied preferences for SWB by administering a survey that posed a series of binary, hypothetical choices.[64] (For example, one choice was between a job paying $80,000/year with reasonable hours, permitting seven-and-a-half hours sleep each night, and a job paying

---

61. Laura A. King & Christie K. Napa, *What Makes a Life Good?*, 75 J. PERSONALITY & SOC. PSYCHOL. 156, 158 (1998).

62. *Id.* at 158–61.

63. Matthew Adler & Paul Dolan, *Introducing a "Different Lives" Approach to the Valuation of Health and Well-Being* 12 (Kenan Inst. for Ethics at Duke Univ., Working Paper No. 5, 2012).

64. Benjamin et al., *supra* note 51, at 2087.

$140,000 year allowing only six hours of nightly rest.)[65] Their clever research strategy was to ask respondents, first, which choice would maximize SWB (either happiness or life satisfaction, depending on the sample, via a question such as, "Between these two options, taking all things together, which do you think would give you a happier life as a whole?") and, second, which option they would choose.[66] Divergent answers to the two questions by a given respondent show that she cares about something other than her predicted happiness/life satisfaction. Finally, a subsample of respondents were asked to identify the better option in light of various factors other than the respondent's happiness/life satisfaction—specifically, her *family*'s happiness, her health, romantic life, social life, control over her life, spirituality, social status, fun, her life's "nonboringness," physical comfort, and sense of purpose.[67]

The authors found that, on average "SWB and choice coincide 83 percent of the time in our data,"[68] but that the strength of this relationship varied substantially depending on choice situation, subject population and questionnaire design. Moreover, in a regression analysis, respondent-predicted SWB was by far the strongest predictor of choice, but "sense of purpose, control over life, family happiness, and social status" also had some role.[69]

An earlier, much smaller study comparing hypothetical choice with predictions of happiness—by Amos Tversky and Dale Griffin— found a larger deviation between the two than the Benjamin et al. study.[70]

In a second study, Benjamin, Heffetz, Kimball, and Rees-Jones were able to compare *actual* choice with anticipated SWB.[71] They asked medical students who had just submitted their rankings for the National Resident Matching Program to assign a level of anticipated SWB to each of their top four chosen residencies (both SWB during

    65. *Id.*

    66. *Id.* at 2087–88.

    67. *Id.* at 2097.

    68. *Id.* at 2085.

    69. *Id.*

    70. Amos Tversky & Dale Griffin, *Endowments and Contrast in Judgments of Well-Being*, *in* CHOICES, VALUES, AND FRAMES 709, 722–23 (Daniel Kahneman & Amos Tversky eds., 2000).

    71. Daniel J. Benjamin, Ori Heffetz, Miles S. Kimball & Alex Rees-Jones, *Can Marginal Rates of Substitution Be Inferred from Happiness Data? Evidence from Residency Choices* 9–15, 31 (Johnson Sch. Research Paper Series, No. 8-2013, 2013), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2221538.

the residency itself, and over the remainder of their lives). The students were also asked to characterize each of the top four choices in terms of a number of attributes, such as the residency's prestige, locational desirability, how stressful it would be, its career value, and its desirability to the student's significant other (if any). The authors used statistical techniques to measure the tradeoffs between the attributes, first in terms of choice (how the students ranked programs) and second in terms of anticipated SWB, and found significant differences. If the students were ranking residencies to maximize anticipated SWB, no such differences should have been observed.

In a recent large-scale survey, Paul Dolan and Robert Metcalfe asked individuals which of eight specific life domains "matters most in your life."[72] Notably, the domains included "Mental wellbeing." A large plurality (42 percent) of respondents identified "Personal relationships" as the most important domain, and "Physical health" was also thus identified by a substantially larger percentage (18 percent) than singled out "Mental wellbeing" (7 percent).[73]

Other bodies of research do not directly address the strength of individuals' preferences for nonexperiential items but still have some probative weight on this issue. First, by contrast with the Dolan and Metcalfe survey, some surveys pose an *open-ended* question about the most significant life domains—asking the respondent to list whichever aspects of his life he cares most about or sees as most important. Often, respondents include domains characterized in nonexperiential terms. For example, a large-scale survey asked British citizens to list the "most important areas of their life" and to place them in priority order.[74] The largest fraction of respondents described relationships with friends or family as the single most important area of their lives, followed by "finances/standard-of-living/housing," and then the respondent's health.[75]

---

72.   DOLAN & METCALFE, *supra* note 18, at 11, 20; *see also* Ed Diener & Christie Scollon, Subjective Well-Being Is Desirable, but Not the Summum Bonum 8–9, 20 (July 2, 2003) (unpublished manuscript) (on file with the *Duke Law Journal*) (asking college students in various countries to rate happiness compared to other values such as wealth, love, health, and so forth).

73.   DOLAN & METCALFE, *supra* note 18, at 5, 20 tbl.10.

74.   Ann Bowling & Joy Windsor, *Towards the Good Life: A Population Survey of Dimensions of Quality of Life*, 2 J. HAPPINESS STUD. 55, 61 (2001).

75.   *Id.* at 59, 61, 64.

It is *possible* that someone who uses nonexperiential language to describe his most important domain actually cares only about his experiences. For example, if I say that friendship is utmost, I might mean to say that (1) feeling satisfied by my friendships is utmost. Or, I might be saying that (2) feeling happy is utmost, and I believe that having friends is the surest route to my own happiness. But it seems more plausible that someone who cared mainly about the quality of his own experiences would use explicitly experiential descriptors ("being happy," "feeling good about my life") to identify the highest-priority part of his life.

A second body of research with some probative weight regarding nonexperiential preference arguments focuses on patients' valuations of health states.[76] This research is one portion of the larger quality-adjusted life year (QALY) literature, which asks patients, health care providers, or members of the general population to value health states on a 0 to 1 scale (with 0 meaning a health state no better than death, and 1 meaning perfect health), using time-tradeoff, standard-gamble, or direct-rating questions.[77]

---

76. *See generally* Xavier Badia, Michael Herdman & Paul Kind, *The Influence of Ill-Health Experience on the Valuation of Health*, 13 PHARMACOECONOMICS 687 (1998); Laura J. Damschroder, Brian J. Zikmund-Fisher & Peter A. Ubel, *Considering Adaptation in Preference Elicitations*, 27 HEALTH PSYCHOL. 394 (2008) [hereinafter Damschroder et al., *Considering Adaptation*]; Laura J. Damschroder, Brian J. Zikmund-Fisher & Peter A. Ubel, *The Impact of Considering Adaptation in Health State Valuation*, 61 SOC. SCI. & MED. 267 (2005); Jason Riis et al., *Ignorance of Hedonic Adaptation to Hemodialysis: A Study Using Ecological Momentary Assessment*, 134 J. EXPERIMENTAL PSYCHOL.: GEN. 3 (2005); Dylan Smith et al., *Mispredicting and Misremembering: Patients with Renal Failure Overestimate Improvements in Quality of Life After a Kidney Transplant*, 27 HEALTH PSYCHOL. 653 (2008); Dylan M. Smith, Stephanie L. Brown & Peter A. Ubel, *Mispredictions and Misrecollections: Challenges for Subjective Outcome Measurement*, 30 DISABILITY & REHABILITATION 418 (2008); Dylan M. Smith, Ryan L. Sherriff, Laura Damschroder, George Loewenstein & Peter A. Ubel, *Misremembering Colostomies? Former Patients Give Lower Utility Ratings Than Do Current Patients*, 25 HEALTH PSYCHOL. 688 (2006) [hereinafter Smith et al., *Misremembering Colostomies*]; Peter A. Ubel, George Loewenstein & Christopher Jepson, *Disability and Sunshine: Can Hedonic Predictions Be Improved by Drawing Attention to Focusing Illusions or Emotional Adaptation?*, 11 J. EXPERIMENTAL PSYCHOL: APPLIED 111 (2005); Peter A. Ubel et al., *Do Nonpatients Underestimate the Quality of Life Associated with Chronic Health Conditions Because of a Focusing Illusion?*, 21 MED. DECISION MAKING 190 (2001); Peter A. Ubel, George Loewenstein, Norbert Schwarz & Dylan Smith, *Misimagining the Unimaginable: The Disability Paradox and Health Care Decision Making*, 24 HEALTH PSYCHOL. S57 (2005) [hereinafter Ubel et al., *Misimagining the Unimaginable*]; Peter A. Ubel, George Loewenstein & Christopher Jepson, *Whose Quality of Life? A Commentary Exploring Discrepancies Between Health State Evaluations of Patients and the General Public*, 12 QUALITY LIFE RES. 599 (2003) [hereinafter Ubel et al., *Whose Quality of Life?*].

77. *See* Matthew D. Adler, *QALYs and Policy Evaluation: A New Perspective*, 6 YALE J. HEALTH POL'Y L. & ETHICS 1, 1 & n.1, 8–9 nn. 25–26 (2006) (collecting sources).

The literature on the hedonic effects of disease documents substantial, sometimes complete hedonic adaptation. "[N]umerous studies have found that people with chronic health conditions as severe as kidney failure or paraplegia report moods that are relatively close to those reported by healthy persons . . . ."[78] Still, healthy individuals frequently assign low QALY values to disease states. This itself does *not* show that healthy individuals have an intrinsic preference not to be diseased. Perhaps, instead, they care mainly about affects and feelings but underestimate the extent of hedonic adaptation to disease.

More interesting, here, is the fact that patients *themselves* tend to give QALY values lower than 1 to their diseases. Several general findings emerge from research about patient valuations. First, these valuations are *higher* than healthy individuals'.[79] Second, these valuations are still substantially below the top of the scale: "both patients and healthy people agree that living with a chronic health condition is worse than living in perfect health . . . ."[80]

Because patients (unlike members of the general population) *are* aware of the moods and feelings associated with their disease, a QALY rating below 1 does suggest an intrinsic preference for health.[81] For example, in one striking study, Dylan Smith and his co-authors asked both individuals with colostomies and former colostomy patients whose bowel function had subsequently been restored to (1) quantify their current moods and life satisfaction using standard SWB questions, and (2) assign a QALY value to having a colostomy, using a time-tradeoff question (which asks about willingness to reduce life span in return for a health improvement).[82] Current patients reported, on average, a degree of life satisfaction only slightly lower than that of former patients and *better* moods. In other words, hedonic adaptation to having a colostomy seems to be virtually complete. However, current patients also, on average, expressed a willingness to reduce their life span by 16 percent in exchange for a return to perfect health.[83] This seems to be substantial

---

78. Loewenstein & Ubel, *supra* note 19, at 1799.

79. This finding has been termed the "disability paradox." Ubel et al., *Misimagining the Unimaginable*, *supra* note 76, at S57.

80. Damschroder et al., *Considering Adaptation*, *supra* note 76, at 394.

81. *See* Loewenstein & Ubel, *supra* note 19, at 1799–1800, 1803.

82. Smith et al., *Misremembering Colostomies*, *supra* note 76, at 689–90.

83. *See id.* at 192 tbl.2. The life-satisfaction and mood differences between current and former patients were not statistically significant. *Id.*

(if not conclusive) evidence that colostomy patients care about healthy bowel function as such, not just about the moods and feelings associated with having or lacking a healthy bowel.[84]

To sum up, the extent to which individuals intrinsically care about nonexperiential aspects of their lives warrants much more intensive study. The existing body of empirical work bearing on this issue is quite small and hardly suffices to demonstrate that individuals generally lack such preferences. Policy tools for implementing a preference-based account of well-being should therefore be structured to allow for the *possibility* that some individuals, at least, have preferences with nonexperiential fundamental arguments.

C. *The Policy Relevance of Subjective Well-Being Surveys: Two Defenses*

We can now distinguish two possible defenses of the policy relevance of SWB surveys. By SWB survey, again, I mean a survey that asks for a numerical rating of the respondent's overall or momentary happiness, life satisfaction, or some other aspect of her mental life.

One defense, the PR (preference-realization) defense, adopts a preference view of well-being—analyzing an individual's well-being in terms of the realization of her preferences, and allowing that the fundamental arguments for those preferences can include items other than her mental states (such as her health, liberty, relationships, accomplishments, and so forth). The content of individual i's ranking of possible outcomes, R_i, is for her to determine; her self-rated SWB, in turn, is taken as a defeasible indicator of whether the actual outcome is located high or low in this ranking.

Why would this be the case? Although preference realization and the preference-holder's feelings of satisfaction are distinct, there is some plausibility in thinking that an individual's answer to an SWB survey is *good if not perfect evidence* of her preference realization. After all (so the account goes) the typical arguments for preferences, if not necessarily mental states, *are* still features of the world to which individuals have good epistemic access. Some proportion of the population will be hypochondriacs, some proportion will be in denial about their diseases, but most people will know pretty well how
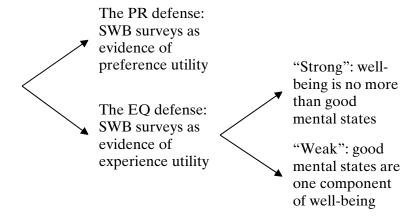
---

84. It is not conclusive because colostomy patients may be mispredicting the hedonic benefit of a return to perfect health.

healthy they are. Some poor souls *will* be deceived by spouses, friends, or colleagues, but many others will have an accurate sense of the quality of their social lives.

A quite different defense, the EQ (experience-quality) defense, sees an appropriately designed SWB survey as evidence of the quality of an individual's mental states. For example, an individual who says her happiness is 7 on a 1 to 7 scale gives us strong evidence that her current mood and emotions are positive and that she is not in pain. These facts about her mental life (specifically, her affective state) contribute favorably to her well-being. Or, an individual who says she is "not satisfied" indicates that she is currently feeling unsatisfied—and, plausibly, feelings of satisfaction or dissatisfaction also are mental facts relevant to well-being. The EQ defense might be allied to an experientialist account of well-being, and thus take the strong form. Less ambitiously, the EQ defense might reject or bracket the experientialism requirement, and instead claim only that good mental states are *one* dimension of well-being, and that SWB surveys in turn provide evidence regarding this dimension.

More simply, this duality of roles for SWB surveys can be expressed using Kahneman's bifurcation between decision utility (better, preference utility) and experience utility.[85] An individual's answer to an SWB survey might be evidence of her preference utility. Alternatively, it might be evidence of her experience utility.

Figure 1. Possible Defenses of the Policy Relevance of Subjective Well-Being Surveys

The PR defense: SWB surveys as evidence of preference utility

The EQ defense: SWB surveys as evidence of experience utility

"Strong": well-being is no more than good mental states

"Weak": good mental states are one component of well-being

---

85.   *See supra* note 56 and accompanying text.

In the remainder of this Article, I carefully work through this flowchart.

Some SWB scholars will be impatient with close consideration of the PR account. "This is just a straw man," they might say. "The point of SWB surveys is *obviously* to measure experience utility, not preference utility."

However, a careful reading of SWB scholarship suggests that some prominent SWB researchers *do* adopt the PR defense. For example, Andrew Clark, Michael Frijters, and Paul Shields, in a wide-ranging article on the Easterlin paradox, explain that they see this as a paradox concerning *decision* utility. The flat time trend of SWB scores suggests (as they see it) that increasing per-capita GDP makes little difference to *decision* utility over time within a given country.[86] They propose to "explain" the paradox by constructing a *decision* utility function that includes relative as well as absolute income as arguments and that incorporates adaptation effects. As they explain:

> The explanation of the Easterlin paradox detailed in this paper rests on the ways in which income translates into utility. It is important to be clear about the logical step that we are taking here. While the paradox is couched in terms of income and happiness, we are going to appeal to a specific type of utility function to account for it. In other words, we imagine that *happiness scores provide information about utility.* . . . Section 4 will then explicitly set out the evidence linking happiness and utility.[87]

And what they then write at the beginning of Section 4 is: "In this section we ask what basis there is for believing that happiness is a reasonable measure of the economic notion of (*decision*) utility, i.e., the thing whose maximization leads to choice behavior."[88]

---

86. Andrew E. Clark, Paul Frijters & Michael A. Shields, *Relative Income, Happiness, and Utility: An Explanation for the Easterlin Paradox and Other Puzzles*, 46 J. ECON. LITERATURE 95, 99–106, 115 (2008). The "paradox," first observed by Richard Easterlin, is the substantial growth of income in many countries without a corresponding rise in happiness levels. *Id.* at 95. *But see* Daniel W. Sacks, Betsey Stevenson & Justin Wolfers, *Subjective Wellbeing, Income, Economic Development and Growth*, *in* . . . AND THE PURSUIT OF HAPPINESS: WELL-BEING AND THE ROLE OF GOVERNMENT 59, 59 (Philip Booth ed., 2012) (challenging the existence of the paradox).

87. Clark et al., *supra* note 86, at 99 (emphasis added).

88. *Id.* at 115 (emphasis added). At the conclusion of Section 4 of the article, Clark et al. note reason for caution about "the link between happiness and utility" and here, again, make explicit that by "utility" they mean "decision utility." *Id.* at 121; *see also* Erik Angner, Subjective Well-Being: When, and Why, It Matters 18 (Aug. 31, 2012) (unpublished manuscript), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2157140 (suggesting that a happiness

Others defend SWB surveys as evidence of "utility" without disambiguating that term.[89] Indeed, the failure of the literature to draw a sharp distinction between SWB surveys as evidence of preference versus experience utility—notwithstanding Kahneman's terminological efforts—is one aspect of the conceptual fuzziness of this literature. The statement that SWB surveys are policy-relevant because they indicate "utility" invites the careless reader to commit a fallacy of equivocation—hybridizing the strong evidentiary role of SWB surveys qua *experience* utility, and the welfare relevance of preference realization. The equivocation runs as follows: "SWB surveys provide powerful evidence of (experience) utility. But of course (preference) utility is policy-relevant, on a preference account of well-being. Thus SWB surveys are policy-relevant."

It is therefore important to separate out the PR and EQ accounts—more specifically, to show why the PR account is problematic (Part II of this Article), before turning to the more plausible view that SWB surveys are useful evidence to governments in providing information about the quality of individuals' mental states.

II. DO SUBJECTIVE WELL-BEING SURVEYS FURNISH GOOD EVIDENCE OF PREFERENCE UTILITY?

This Part critically evaluates the possible use of SWB scores as evidence of preference utility. Section A outlines some significant obstacles to such inference: *scale recalibration*, *preference heterogeneity*, *evaluation error*, *and miscommunication*. Section B discusses the extent to which these potential obstacles can be circumvented via econometric techniques.

To be sure, SWB surveys might be imperfect evidence of preference utility but still *better* evidence than other sources. Section C argues to the contrary. Stated-preference surveys *dominate* SWB surveys *as evidence of preference utility*. *If* well-being reduces to

---

measure can provide useful information about preference-realization insofar as individuals want happiness or happiness is sufficiently correlated with preference-realization); Peter Railton, Subjective Well-Being as Information and Guidance 33 (Mar. 2012) (unpublished manuscript), *available at* http://profron.net/happiness/files/readings/Railton_SubjectiveWellBeing.pdf (suggesting that the affective and life-satisfaction components of an individual's SWB provide information about her success in attaining, respectively, short-term and long-term goals).

89.   *See, e.g.*, Blanchflower & Oswald, *supra* note 24, at 1361; Rafael Di Tella & Robert MacCulloch, *Some Uses of Happiness Data in Economics*, J. ECON. PERSP., Winter 2006, at 25, 28.

preference realization, and *if* surveys (in addition to behavioral data) provide useful information about what individuals prefer, SWB surveys have no advantages over stated-preference surveys and many disadvantages.

In much of this Part, I focus on the proposal to use SWB surveys to calculate monetary equivalents for nonmarket goods. On the PR account of SWB surveys, this proposal says that SWB surveys are useful information regarding the amount that individuals are *willing to pay/accept* for those goods. It will be helpful, in assessing the value of SWB surveys, to do so with reference to a specific policy tool that such surveys might be seen as informing. And the monetary-equivalent proposal is one of the key recommendations to emerge from the SWB literature. Moreover (as will be explained below) this proposal is less informationally demanding than proposals that require the summation or averaging of SWB scores (for example, the calculation of gross national happiness). The use of SWB surveys to calculate willingness-to-pay/accept (WTP/WTA) amounts requires only that the surveys evidence individuals' *ordinal* preference utility. Thus, if they perform poorly in this ordinal role—as I will contend— then a fortiori they are problematic as evidence of individuals' *cardinal* preference utilities.[90]

This Part generally focuses on life-satisfaction rather than happiness questions. Asking an individual to quantify how satisfied she is with her life would seem to be a *relatively* more promising vehicle for ascertaining her degree of preference utility than asking her how happy she is. If life-satisfaction questions are, in fact, poor evidence of preference utility—as I will contend—then a fortiori happiness questions are as well.

Finally, so as to simplify the exposition, I will assume that each individual's preference ranking of outcomes is solely a function of her own attributes in the outcomes. Individual *i* prefers outcome *x* to outcome *x** just in case she prefers *A* to *A**, where *A* are her attributes in *x* and *A** in *x**.[91] "Attributes" is understood in a broad

---

90. *See infra* Part II.A.3.

91. To say that a given individual *i* prefers bundle *A* to bundle *A** is a shorthand for saying: *i* prefers it to be the case that *she* has attributes *A*, as opposed to it being the case that *she* has attributes *A**. Thus, when we compare how various individuals rank a given set of attribute bundles, we are not comparing how those individuals rank the very same states of affairs. Rather, we are comparing how one individual ranks states of affairs specified as those in which *she* has various possible attributes, to how a second individual ranks states of affairs specified as those in which *he* has various possible attributes.

sense to include nonmental and relational attributes. An individual's health is an attribute, as is the quality of public goods she enjoys, or the happiness of her children. This ranking is representable by a preference-utility function $u_i(.)$, such that i prefers attribute bundle A to A^* if and only if $u_i(A) > u_i(A^*)$.

A. *Inferring Preference Utility from Life-Satisfaction Questions: Some Obstacles*

1. *Scale Recalibration and Preference Heterogeneity.* Preference heterogeneity and scale recalibration (heterogeneity in the utility scales used to express preferences) are conceptually distinct obstacles to inferring preferences from life-satisfaction surveys, but for expository purposes are discussed in tandem here.

Imagine that Phyllis has a current income of y and a level z of some nonincome attribute (an environmental good, health, etc.). Asked how satisfied she is with her life, she says "7." Gina has a smaller current income of y^*, which is less than y by amount Δy. Her level of the other attribute is z^*, which exceeds z by amount Δz. When asked how satisfied *she* is with her life, Gina also says "7."

Assume, now, that the following are true. (1) Phyllis and Gina have *identical* preferences. Each has the very same ranking of attribute packages (here, combinations of income and the nonincome attribute). For every A and A^*, either both individuals prefer A to A^*, or both prefer A^* to A, or both are indifferent. (2) Each uses the same numerical *scale* to express her ranking of attribute packages. In other words, Phyllis articulates her ranking via a preference-utility function $u_{Phyllis}(.)$, such that $u_{Phyllis}(A) > u_{Phyllis}(A^*)$ whenever Phyllis prefers A to A^*. And—it turns out—the preference-utility function $u_{Gina}(.)$ by means of which Gina expresses *her* ranking of the bundles is exactly the same function as $u_{Phyllis}(.)$. For every bundle A, $u_{Phyllis}(A) = u_{Gina}(A)$.

Finally, (3) Gina and Phyllis each respond to a life-satisfaction question by articulating her *preference utility* for her current attribute package. When Gina possesses attribute bundle A and is asked "How satisfied are you with your life?," the answer she gives is just $u_{Gina}(A)$. Similarly, if Phyllis possesses bundle B and is asked the same question, the answer she gives is $u_{Phyllis}(B)$.

If all of the premises just mentioned hold true, we can make inferences about Phyllis and Gina's WTP/WTA amounts for the nonincome attribute. For example, in the case described three

paragraphs above, we can infer that Phyllis, Gina, and anyone else with the same preferences as them is willing to pay Δy dollars for an increase in the z attribute by amount Δz. If the case were varied, so that Gina's answer to the life-satisfaction question is a number greater than 7, we could infer that Phyllis, Gina, and anyone with the same preferences is willing to pay more than Δy for an increase in the z attribute by amount Δz. Finally, an answer less than 7 would indicate that Phyllis, Gina, and anyone with the same preferences is willing to pay less than Δy for that increase in the z attribute.

But consider, now, relaxing the premise of identical scales. As economists are very well-aware, a utility function representing an ordering of attribute bundles (or any other items) is hardly unique; rather, it is unique only up to an ordinal ("increasing") transformation.[92] Even though Phyllis and Gina have the same ranking of possible bundles, it need *not* be the case that $u_{Phyllis}(.)$—the mathematical function which Phyllis uses to express that ranking—is the same as $u_{Gina}(.)$. For example, if there are six bundles, and each of the women prefers the first to the second, the second to the third, and so forth, then Phyllis might express this preference via the numbers 9, 8, 7, 6, 5, 4, whereas Gina might use the numbers 6, 5.5, 4, 3.5, 3, 2.5.

Even holding fixed the assumption of identical preferences, scale heterogeneity interferes with our ability to make inferences regarding Phyllis's and Gina's WTP/WTA amounts. Back to the case in which Phyllis with attributes (y, z) says that her life satisfaction is 7, and Gina with attributes $(y - \Delta y, z + \Delta z)$ also says that her life satisfaction is 7. What can we infer about how much Phyllis and Gina are willing to pay for an increase in the z attribute by Δz? Not much. It *might* be the case that they are willing to pay exactly Δy for Δz. But it *might* also be that they are willing to pay less than Δy. (Imagine that both Phyllis and Gina prefer the package (y, z) to the package $(y - \Delta y, z + \Delta z)$, but Phyllis represents this preference by assigning the number 7 to the first package and 4 to the second, whereas Gina represents this preference by assigning the number 9 to the first package and 7 to the second.) Finally, it *might* be the case that they are willing to pay *more* than Δy.

Preference heterogeneity further complicates the picture. If we allow for the possibility that Phyllis and Gina may have *different*

---

    92. Let $u(.)$ be one function of attribute bundles, $u*(.)$ a second. To say that $u*(.)$ is an increasing transformation of $u(.)$ means that whenever $u(A) = u(B)$, $u*(A) = u*(B)$ and whenever $u(A) > u(B)$, $u*(A) > u*(B)$.

rankings of the attribute bundles, the fact that Phyllis has told us her preference utility for *one* bundle (y, z), and Gina has told us her preference utility for *one* bundle $(y - \Delta y, z + \Delta z)$, permits no inferences about whether Phyllis is willing to pay more, less, or exactly Δy for Δz. Nor does it permit inferences about whether Gina is willing to pay more, less, or exactly Δy for Δz. To make inferences about Phyllis's WTP amount—absent an assumption of preference homogeneity—we would need to hear Phyllis express her preference utility for both bundles (and similarly for Gina).

To see this in a yet simpler way: imagine that Phyllis is asked about her utility for chocolate ice cream and says "7," whereas Gina is asked about her utility for vanilla ice cream and also says "7." Unless we have some reason to believe that the two have the same tastes in ice cream (and also are using the same utility scale), can we conclude anything about whether Phyllis prefers chocolate, prefers vanilla, or is indifferent? Can we conclude anything about whether Gina prefers chocolate, prefers vanilla, or is indifferent? Of course not.

Nor do problems of scale and preference heterogeneity disappear when we have *multiple* answers to life-satisfaction questions from the same person, as with longitudinal (panel) data. Vary the case under discussion so that Howard at Time 1, when he has income y and nonincome attribute z, says that his life satisfaction is 7. At Time 2, when Howard's income has decreased to $y - \Delta y$ and his level of the nonincome attribute has increased to $z + \Delta z$, he *also* says that his life satisfaction is 7. Can we infer that Howard's WTP for Δz is Δy? Not necessarily. One possibility is that Howard has the same ranking of attribute bundles at each time but uses a different preference-utility function to express that ranking at Time 2 than at Time 1.

It is also possible that the ranking of bundles has itself shifted. There are various potential systematic sources of intrapersonal preference heterogeneity. For example, individuals' preferences may tend to change as they age. A different kind of intrapersonal preference heterogeneity arises from adaptive/counteradaptive preferences. Imagine that an individual's ranking of attribute bundles at a particular point in time varies, in a systematic way, with his past levels of the attributes. One possibility—adaptive preferences— would be that higher levels of an attribute in the recent past tend to induce the individual to have a weaker preference for the attribute. For example, if the individual has a current income of $100,000 and

has had that income for several years, he may require *more* immediate income compensation for a given change in his health state, than if he has a current income of $100,000 but in prior years had an income of $50,000.[93] Reciprocally, if higher levels of an attribute in the recent past induce a stronger preference for the attribute, the individual's preferences would be counteradaptive.

In short, inter- and intrapersonal preference heterogeneity and scale heterogeneity are—in principle—obstacles to inferring preferences from SWB surveys. What is the evidence that these phenomena actually occur?

The empirical literature on preferences provides ample evidence of interpersonal preference heterogeneity. Numerous studies using standard preference data, other than SWB studies themselves (that is, behavioral data or stated-preference data), in various contexts, have confirmed the common-sense point that different individuals often have different rankings of commodity bundles, income-leisure bundles, different degrees of risk aversion, and so forth.[94]

Moreover, various literatures document interpersonal "scale recalibration": interpersonal heterogeneity in the numerical scales that individuals use to report various phenomena. For instance, a substantial literature documents interpersonal heterogeneity in health rating scales.[95] Here are a few illustrative examples. In one study, Peter Ubel and his co-authors asked respondents to rate their own health on a scale from 0 to 100. One group was told that 100

---

93. For a formal model of adaptive preferences, see Clark et al., *supra* note 86, at 104–06.

94. *See* ADLER, *supra* note 27, at 279 n.57 (collecting sources).

95. *E.g.*, Wim Groot, *Adaptation and Scale of Reference Bias in Self-Assessments of Quality of Life*, 19 J. HEALTH ECON. 403 (2000); Heather P. Lacey et al., *Are They Really That Happy? Exploring Scale Recalibration in Estimates of Well-Being*, 27 HEALTH PSYCHOL. 669 (2008); Maarten Lindeboom & Eddy van Doorslaer, *Cut-Point Shift and Index Shift in Self-Reported Health*, 23 J. HEALTH ECON. 1083 (2004); Erik Meijer, Arie Kapteyn & Tatiana Andreyeva, *Internationally Comparable Health Indices*, 20 HEALTH ECON. 600 (2011); Debby Postulart & Eddy M.M. Adang, *Response Shift and Adaptation in Chronically Ill Patients*, 20 MED. DECISION MAKING 186 (2000); Joshua A. Salomon, Ajay Tandon & Christopher J.L. Murray, *Comparability of Self Rated Health: Cross Sectional Multi-Country Survey Using Anchoring Vignettes*, 328 BRIT. MED. J. 258 (2004), doi:10.1136/bmj.37963.691632.44; Amir Shmueli, *Reporting Heterogeneity in the Measurement of Health and Health-Related Quality of Life*, 20 PHARMACOECONOMICS 405 (2002); Amir Shmueli, *Socio-Economic and Demographic Variation in Health and in Its Measures: The Issue of Reporting Heterogeneity*, 57 SOC. SCI. & MED. 125 (2003); Mirjam A.G. Sprangers & Carolyn E. Schwartz, *Integrating Response Shift into Health-Related Quality of Life Research: A Theoretical Model*, 48 SOC. SCI. & MED. 1507 (1999); Peter A. Ubel, Aleksandra Jankovic, Dylan Smith, Kenneth M. Langa & Angela Fagerlin, *What Is Perfect Health to an 85-Year-Old?: Evidence for Scale Recalibration in Subjective Health Ratings*, 43 MED. CARE 1054 (2005).

represented "perfect health"; a second, that it represented "perfect health for someone your age"; a third, that it represented "perfect health for a 20-year-old."[96] The third scale was anchored on a specific age, whereas the second was anchored on the respondent's age. Average ratings using the first two scales were very close. By contrast, average ratings using the third were lower—suggesting that respondents interpreted "perfect health" as "perfect health for someone your age," a kind of recalibration.[97]

In another study, Joshua Salomon and his co-authors presented respondents in different countries with "mobility vignettes": descriptions of hypothetical subjects, highlighting their ability to move around, using language such as the following: "Rina has had a stiff neck for the last 10 days and it makes her move around slowly as any sudden movement causes pain." Or, "Louis is able to move his arms and legs, but requires assistance in standing up from a chair or walking around the house. Any bending is painful, and lifting is impossible."[98] Respondents were asked to describe the subject's degree of mobility, using five ordered categories: no, mild, moderate, severe, or extreme difficulty moving around. Respondents tended to order the vignettes similarly, but to vary by country and age in their application of the five categories, suggesting scale heterogeneity. For example, both Chinese and Sri Lankan respondents might tend to place Louis in a less mobile category than Rina; but the plurality of Sri Lankans might rate Louis as having "extreme" difficulty with respect to mobility, whereas the plurality of Chinese might rate his difficulty as "severe."[99]

As already mentioned, individuals with a particular disease tend to assign higher QALY values to that disease than do members of the general population.[100] There are at least three different explanations for this divergence: (1) patients and members of the general population have different factual understandings of the impact of the disease state on the patient's life (for example, its hedonic impact); (2) patients and members of the general population have different preferences regarding disease states; and (3) patients and members of the general population use different rating scales to express their

---

96.   Ubel et al., *supra* note 95, at 1055 tbl.1.

97.   *Id.* at 1056.

98.   Salomon et al., *supra* note 95, at 2.

99.   *Id.* at 3–4.

100. *See supra* notes 76–84 and accompanying text.

preferences (scale recalibration).[101] The relative importance of these three explanations is contested, but substantial evidence suggests that scale recalibration is at least one part of the picture.[102]

Heterogeneity in health rating scales may furnish *direct* evidence of heterogeneity in preference-utility scales. In some contexts, individuals are using the health number to express their preferences over health states. At a minimum, heterogeneity in the scales used to assign numbers to health states or other phenomena provides *circumstantial* evidence of scale recalibration with respect to preference utility. If individuals vary in how they use numbers to rate health, height,[103] political efficacy, or visual acuity,[104] then (absent evidence to the contrary) we should be worried about variation in preference-utility scales.

Bernard van Praag's survey work using "income evaluation" questions is an important body of research—suggestive either of heterogeneity with respect to individuals' preferences regarding income, or heterogeneity in preference-utility scales. The respondent was asked to state the range of incomes she considered to fall in each of six categories: "very bad," "bad," "insufficient," "sufficient," "good," and "very good."[105] The cutoffs for each category were then correlated with the respondent's characteristics, including her actual income. Van Praag found that an increase in respondent's income tended to shift each range upward. The more the respondent's actual income, the larger a hypothetical income amount had to be before she would categorize it as "very good," "good," "sufficient," and so forth.

My discussion, thus far, has focused on studies that document interpersonal preference heterogeneity and/or scale recalibration via cross-sectional research designs, showing how different people rank

---

101. To be sure, scale recalibration would only be relevant for QALY values elicited via ratings, rather than via time-tradeoff or standard-gamble questions.

102. *See* Ubel et al., *Whose Quality of Life?*, *supra* note 76, at 604–05.

103. *See* Andrew J. Oswald, *On the Curvature of the Reporting Function from Objective Reality to Subjective Feelings*, 100 ECON. LETTERS 369, 370–71 (2008).

104. *See* Gary King, Christopher J.L. Murray, Joshua A. Salomon & Ajay Tandon, *Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research*, 98 AM. POL. SCI. REV. 191 (2004) (discussing the use of "vignettes" to correct for scale heterogeneity, and illustrating this technique with respect to measures of political efficacy and visual acuity); *see also* Mary Steffel & Daniel M. Oppenheimer, *Happy by What Standard? The Role of Interpersonal and Intrapersonal Comparisons in Ratings of Happiness*, 92 SOC. INDICATORS RES. 69 (2009) (finding scale heterogeneity with respect to happiness scales).

105. *See* BERNARD M.S. VAN PRAAG & ADA FERRER-I-CARBONELL, HAPPINESS QUANTIFIED: A SATISFACTION CALCULUS APPROACH 5–6, 34–42 (2004).

or rate various phenomena. There appears to be less work using longitudinal data, which could directly evidence intrapersonal preference heterogeneity or scale recalibration. However, cross-sectional studies may be suggestive of intrapersonal phenomena. For example, a cross-sectional finding that older respondents use a different scale than younger respondents to rate their health suggests that a given individual will tend to change her health-rating scale as she ages. Van Praag's research on income evaluation questions is some evidence of adaptive preferences with respect to income.

In sum, more research is needed, but existing empirical work provides evidence of preference heterogeneity and heterogeneity in preference-utility scales. These are empirically genuine—not merely theoretical—obstacles to using SWB surveys to infer preference utility.

2. *Evaluation Error and Miscommunication*. Return again to the case in which Gina, while in possession of a particular attribute bundle, quantifies her life satisfaction as a particular number, and Phyllis, while in possession of a different bundle, also quantifies her life satisfaction as that same number. As discussed, if Gina and Phyllis (1) have the same preferences, represented via (2) the same preference-utility function, *and* in addition (3) the answer each gives to a life-satisfaction survey is exactly equal to her preference utility for her current attributes, then we can infer their common WTP/WTA amounts.

"Evaluation error" and "miscommunication" constitute different types of failure of this last condition. The respondent may *misapply* the utility function to her actual bundle of attributes: this is "evaluation error." Although her actual attributes are bundle B, Gina incorrectly perceives her current preference utility to be some value *other than* $u_{Gina}(B)$. When asked, "How satisfied are you with your life?," Gina articulates that value, not $u_{Gina}(B)$.

Alternatively, Gina's perceived preference utility may be correct (she has attributes B, and indeed perceives her preference utility to be $u_{Gina}(B)$); but she responds to the question "How satisfied are you with your life?" by articulating some value other than her perceived preference utility.

A 1999 book chapter by Norbert Schwarz and Fritz Strack, reviewing research on the psychology of life-satisfaction questions,

DUKE LAW JOURNAL   [Vol. 62:1509

provides ample evidence of both of these effects.[106] Consider, first, evaluation error. This can arise in various ways. The respondent's preferences might depend upon features of outcomes about which she has imperfect information[107] (as in the example of a preference for a faithful spouse). Or she might possess the relevant information, but fail to access it. Schwarz and Strack show how the latter occurs: the survey instrument, the respondent's mood, or recent events may focus her attention on certain information about her current attributes and divert attention from other data.

> When asked, "Taking all things together, how would you say things are these days?" respondents are ideally assumed to review the myriad of relevant aspects of their lives and to integrate them into a mental representation of their life as a whole. In reality, however, individuals rarely retrieve all information that may be relevant to a judgment. Instead, they truncate the search process as soon as enough information has come to mind to form a judgment with sufficient subjective certainty. Hence, the judgment is based on the information that is most *accessible* at that point in time. In general, the accessibility of information depends on the recency and frequency of its use.[108]

For example, researchers found a strong correlation between respondent's dating frequency and life satisfaction only when the question about dating frequency preceded the life-satisfaction question.[109]

Schwarz and Strack also identify much evidence of mood effects.

> [J]udgments of well-being are a function not only of what one thinks about but also of how one *feels* at the time of judgment. A wide range of experimental data confirms this intuition. Finding a dime on a copy machine, spending time in a pleasant rather than an

---

106. Norbert Schwarz & Fritz Strack, *Reports of Subjective Well-Being: Judgmental Processes and Their Methodological Implications*, *in* WELL-BEING: THE FOUNDATIONS OF HEDONIC PSYCHOLOGY, *supra* note 1, at 61. The terms "evaluation error" and "miscommunication" are my own; Schwarz and Strack do not use these terms.

107. *Cf.* Hendrik Jürges, *Unemployment, Life Satisfaction and Retrospective Error*, 170 J. ROYAL STAT. SOC'Y 43, 44 (2007) (discussing the effect of inaccurate memory on survey responses).

108. Schwarz & Strack, *supra* note 106, at 63 (citations omitted). A recent review concludes: "What information people attend to when responding to surveys can strongly affect life satisfaction judgments." Ed Diener, Ronald Inglehart & Louis Tay, *Theory and Validity of Life Satisfaction Scales*, SOC. INDICATORS RES. 11 (2012), link.springer.com/article/10.1007 %2Fs11205-012-0076-y?LI=true.

109. Schwarz & Strack, *supra* note 106, at 63.

unpleasant room, or watching the German soccer team win rather than lose a championship game all resulted in increased reports of happiness and satisfaction with one's life as a whole.[110]

Mood effects might involve a kind of evaluation error: mood makes salient or less visible "mood-congruent" or incongruent information, respectively. (In plainer English: happy people tend to see the ways in which their lives are fulfilling their preferences, less happy people the ways in which they are not.) Mood effects might instead fall under the heading of miscommunication: the respondent might understand a life-satisfaction question as asking for a quantitative measure of her hedonic states—for *experience* utility—rather than for her preference utility.

"Miscommunication" is a broad category: anything that leads the respondent to articulate some number *other than* her perceived preference utility for her current attributes is a kind of miscommunication. For example, cultural norms may encourage individuals with especially high or low levels of preference utility to communicate a more mediocre number (or, conversely, push individuals whose preferences are only modestly realized to claim greater success).[111] Depending on the order of questions, conversational norms may induce the respondent to articulate her "domain satisfaction"—subutility for a subset of attributes—rather than her preference utility for the totality of her attributes.[112]

3. *A Note on Cardinality*. Much work in economics distinguishes between ordinal and cardinal utility. For purposes of the present discussion, the distinction can be framed as follows.

An individual's *ordinal* preference-utility function captures her ranking of attribute bundles. The function $u_{Gina}(.)$ is an ordinal utility function for Gina if, whenever Gina ranks one bundle over a second bundle, this function assigns the first a higher number. As already discussed, an ordinal utility function is not unique, but merely unique up to an increasing transformation. If $u_{Gina}(.)$ is an ordinal utility

---

110. *Id.* at 74 (citations omitted). Subsequent findings with respect to weather effects have been mixed. *See* Diener et al., *supra* note 108, at 18–19; Sylvia Kämpfer & Michael Mutz, *On the Sunny Side of Life: Sunshine Effects on Life Satisfaction*, 110 SOC. INDICATORS RES. 579 (2013).

111. On cultural differences in responses to SWB surveys, see generally Shigehiro Oishi & Ed Diener, *Culture and Well-Being: The Cycle of Action, Evaluation, and Decision*, 29 PERSONALITY & SOC. PSYCHOL. 939 (2003).

112. *See* Schwarz & Strack, *supra* note 106, at 64.

function for Gina, and $u^*(.)$ an increasing transformation of $u_{Gina}(.)$, then $u^*(.)$ is also an ordinal utility function for Gina.

An individual's *cardinal* preference-utility function not only captures her ranking of attribute bundles, but also captures some other feature of her preferences (for example, her ranking of *lotteries* over bundles, or her time-tradeoff preferences), such that the utility numbers assigned to bundles can meaningfully be added together.[113]

Using SWB surveys to infer preference utility so as to calculate WTP/WTA amounts rests upon the various assumptions surveyed in this Section, namely, that the individuals surveyed have the same preferences, that their answers to the SWB survey are not corrupted by evaluation error or miscommunication, and that these individuals express their preferences using the same preference-utility function. But using surveys to infer WTP/WTA amounts does not rest upon the *further* assumption that this common utility function is cardinal.

To see why not, assume that Phyllis and Gina have the same preferences (the same ranking of attribute bundles) and express this ranking using the very same utility function $u(.)$. Phyllis says that her life satisfaction with bundle (y, z) is 7 because $u(y, z)$ is 7. Gina says her life satisfaction with bundle $(y - \Delta y, z + \Delta z)$ is 7 because $u(y - \Delta y, z + \Delta z)$ is 7. Then we can correctly infer that Phyllis and Gina are each willing to pay exactly Δy for Δz. And—here's the critical point—we could infer the same if Phyllis and Gina were using $u^*(.)$ rather than $u(.)$ to express their common bundle rankings, where $u^*(.)$ is any increasing transformation of $u(.)$.

---

113.   Defining cardinality is a subtle matter, and the definition offered here is rough, but will suffice for the purpose of this Article. Assume that *i* has a preference structure represented by $v(.)$ and by any other function $v^*(.)$ just in case $v^*(.)$ is an *eligible transformation* of $v(.)$. For example, it is well known that, in the case of preferences regarding lotteries, $v^*(.)$ needs to be a positive linear transformation of $v(.)$, that is, $v^*(.) = av(.) + b$, with *a* positive. Similarly, in the case of time-tradeoff preferences, $v^*(.)$ may need to be a positive ratio transformation of $v(.)$, that is, $v^*(.) = av(.)$, with *a* positive.

        Now consider two possible series of attribute packages: *A*, *B*, … versus *A'*, *B'*, …. Utility function $v(.)$ is cardinal if it represents a feature of preferences such that, for every eligible $v^*(.)$, $v(A) + v(B) + … \geq v(A') + v(B') + …$ if and only if $v^*(A) + v^*(B) + … \geq v^*(A') + v^*(B') + …$. Utility function $v(.)$ and all eligible transformations thereof assign overall sums to series of packages so as to rank these series the same way. It is therefore "meaningful" to engage in an operation such as $v(A) + v(B) + …$, and so $v(.)$ can be termed "cardinal."

        Note that, if $v^*(.)$ is a positive ratio transformation, $v(.)$ is cardinal. Moreover, if all series being compared have the same number of terms, and $v^*(.)$ is a positive linear transformation, $v(.)$ is cardinal. Thus "cardinality," in the scholarly literature, is often associated with being a ratio and/or linear transformation.

However, some plausible uses of SWB data do require cardinal utility. Consider, in particular, the proposal to calculate gross/average happiness of various groups, either entire nations or subnational groups. For purposes of this Part, let us construe this as the proposal to use SWB surveys so as to estimate the gross/average *preference* utility of various groups.

Such a proposal imposes an additional demand upon such surveys: namely, that respondents have the same preferences, be uncorrupted by evaluation error or miscommunication, and respond to the survey by applying a common *cardinal* utility function to their current attributes. For example, imagine that there are bundles A, B, C, and D. Phyllis and Gina each prefer A to B to C to D. They also have other common preferences (for example, a common ranking of attribute lotteries, or common time-tradeoff preferences), accurately represented by a cardinal preference-utility function $v(.)$, which assigns the bundles the values $v(A) = 20$, $v(B) = 15$, $v(C) = 10$, $v(D) = 0$.

Imagine that, in 2010, Phyllis has bundle A and Gina has bundle D. In 2011, Phyllis has bundle B and Gina has bundle C. This means that the two individuals' average preference utility has increased over time, from $(20+0)/2$ to $(15+10)/2$.

Imagine, now, that we try to *infer* the time trend in their average preference utility by looking at the time trend in their average numerical responses to *SWB* surveys. If Gina and Phyllis answer those surveys by articulating preference-utility values that are ordinal but not cardinal, our inference may be incorrect. Consider the ordinal (but not cardinal) utility function $u(.)$, such that $u(A) = 7$, $u(B) = 5$, $u(C) = 4$, $u(D) = 3$. If Gina and Phyllis each use *this* common function in responding to the question "How satisfied are you with your life?," asked in 2010 and 2011, their average SWB score decreases.

In short, anyone proposing to use SWB scores to engage in interpersonal aggregation or averaging of preference-utility values must, inter alia, have strong expectations about how respondents understand the SWB-elicitation question. Respondents must interpret the question as asking for their preference utility (rather than as asking for a measure of their current mood, happiness, and so forth), and indeed as asking for their *cardinal* preference utility. It seems wildly speculative that most respondents do in fact interpret a standard SWB question ("How satisfied are you with your life?") as asking for their cardinal preference utility.

B. *Will Econometric Methodology Solve the Problem?*

To what extent can econometric techniques mitigate the difficulties identified in the previous Section?[114] This Section briefly discusses the question. With apologies to some readers, it presupposes a basic knowledge of econometrics.

Assume—to begin—that the pool of respondents has homogeneous preferences. However, they *are* characterized, potentially, by scale heterogeneity, evaluation error, and miscommunication.

The standard approach researchers employ is to estimate the determinants of SWB using ordinary least squares (OLS). The estimating equation is $SWB_{it} = \beta x_{it} + \varepsilon_{it}$, with x_{it} being a vector of individual i's attributes at time t, including income. The ratio of the coefficient in $\hat{\beta}$ for some nonincome good, to the coefficient on income, immediately yields an estimate of individuals' WTP for the good.

The error term in this equation, ε_{it}, serves to handle *certain* kinds of scale heterogeneity, evaluation error, and miscommunication— namely, when these are caused by unobserved factors that are uncorrelated with the *observed* attributes catalogued in x_{it}. For example, random variations in day-to-day weather may deflate or inflate individuals' moods, randomly changing the mix of information about attributes that is salient to individuals. Transient psychological factors may cause an individual to shift upward or downward the preference scale used to express the (common) attribute ranking.[115]

However, the flaw in this strategy—well-recognized by many economists in the SWB literature[116]—is that there may be unobserved

---

114. For helpful discussions of econometric issues in SWB surveys, see Andrew Clark, Fabrice Etilé, Fabien Postel-Vinay, Claudia Senik & Karine Van der Straeten, *Heterogeneity in Reported Well-Being*, 115 ECON. J. C118 (2005); Ada Ferrer-i-Carbonell & Paul Frijters, *How Important Is Methodology for the Estimates of the Determinants of Happiness?*, 114 ECON. J. 641 (2004); Simon Luechinger, *Valuing Air Quality Using the Life Satisfaction Approach*, 119 ECON. J. 482 (2009); Erzo F.P. Luttmer, *Neighbors as Negatives: Relative Earnings and Well-Being*, 120 Q.J. ECON. 963 (2005); and Nattavudh Powdthavee, *How Much Does Money Really Matter? Estimating the Causal Effects of Income on Happiness*, 39 EMPIRICAL ECON. 77 (2010); Dolan et al., *supra* note 11.

115. Blanchflower & Oswald, *supra* note 24, at 1361–62.

116. *See supra* note 114; *see also* Marianne Bertrand & Sendhil Mullainathan, *Do People Mean What They Say? Implications for Subjective Survey Data*, 91 AM. ECON. REV. (PAPERS & PROC.) 67, 67, 71–72 (2001) (arguing against the use of SWB and similar surveys to predict the effect of observable individual attributes on individuals' attitudes because of the correlation of measurement error with those attributes).

individual-specific factors that both cause variation in stated SWB and cause (or otherwise are correlated with) the attributes in x_{it}. For example, the fact that an individual is unusually happy may *both* make her prone to evaluation errors or miscommunications that shift upward the preference-utility scale *and* cause her to earn higher income.

In response to difficulties of this sort, panel data is often used to estimate an OLS equation with individual fixed effects: $\text{SWB}_{it} = \beta x_{it} + f_i + \varepsilon_{it}$. But there are a number of reasons to think that this strategy is *not* a full response to the difficulties under discussion.

One point, a technical one, is that OLS with individual fixed effects assumes a dependent cardinal variable. As discussed earlier, it is far from clear whether an individual in responding to a life-satisfaction question is articulating her perceived cardinal (rather than merely ordinal) utility. Ordinal utility should be estimated via ordered probit or logit—and incorporating fixed effects in these models yields a biased estimate of β.

A second and more substantive worry is that the fixed-effect methodology controls for *time-invariant* sources of scale heterogeneity, evaluation error, or miscommunication. But one theme in the discussion in Section A was that the processes leading to individual SWB responses may change *along with* the change in individual income or other attributes. Intrapersonal scale *recalibration* is just this: an individual with a higher income, health, and so forth, may tend to use a different scale to express his preferences; this effect will show up in ε_{it} rather than in f_i, yielding a biased estimate of β. Similarly, higher levels of certain attributes may cause improvements in individual moods, in turn inducing a systematic shift upward in stated life satisfaction. Finally, cultural norms encouraging respondents to moderate (or inflate) their stated SWB may come into play just when individuals are at higher levels of income or other attributes.

Third, OLS with fixed effects controls for the possibility that time-invariant unobserved factors change the *intercept* of the line associating observed factors with stated life satisfaction, but not the possibility that these skew SWB in more profound ways. For example, individuals with a particular personality trait might be disposed to change the *slope* of the preference-utility function, not just the

intercept.[117] Fourth, OLS with fixed effects has difficulty producing statistically significant estimates for the coefficients on individual observed attributes that do not vary much over time.

Fifth, and this is again a deeper worry, it is hard to see how econometric technique, however sophisticated, can cope with a certain kind of miscommunication effect. OLS with fixed effects (if used to estimate preference utility) starts with the assumption that there is a common preference ranking over attributes, captured by a common utility function with the form $u_{it} = \beta x_{it}$. It then allows for random or individual-specific changes in *expressed* preference utility via f_i and ε_{it}. But if statements of SWB are caused by some feature of individuals' mental states other than their preferences—for example, by their moods—it is puzzling how any such statement can be used to estimate the *preference-utility* function. Consider an analogy. Individuals may have quantitative beliefs of various sorts (to give one example, beliefs regarding the size of the world's population). These beliefs may be caused by various observed and unobserved factors. Would it be sensible to *estimate* the coefficient on these observed factors (the extent to which they change an individual's belief) by asking an individual an SWB question? That would be absurd, because SWB answers are not caused by beliefs about the world's population. But—if SWB answers are indeed expressions of the respondent's hedonic state, not his evaluation of how fully his preferences are realized—why is it less absurd to use SWB answers in estimating preference utility?

Some of these difficulties (although not the last) can be handled via instrumental-variable techniques. But these are used fairly infrequently in the SWB literature, given that valid instruments for income (or other attributes) seem difficult to find in this context.[118] Introducing preference (not merely scale) heterogeneity just further complicates the picture.

C. *The Advantages of Stated-Preference Surveys*

A mere recitation of the various pitfalls in using SWB surveys to estimate preference-utility may ring hollow. The reader may wonder: "Isn't the issue *comparative*? Unless you can identify a *better*

---

117. *See* Clark et al., *supra* note 114, at C119.

118. *See* Powdthavee, *supra* note 114, at 79; Paul Dolan & Robert Metcalfe, *Comparing Willingness-to-Pay and Subjective Well-Being in the Context of Non-Market Goods* 20 (Ctr. for Econ. Performance, Discussion Paper No. 890, 2008).

technique, these pitfalls are unavoidable." But there *is* a better technique: stated-preference surveys.

Stated-preference surveys are widely used in applied economics, particularly to infer individuals' WTP/WTA amounts for nonmarket goods, and also to make other sorts of inferences about preference utility.[119] In the form now favored by the economists who use this methodology, stated-preference surveys ask respondents to rank hypothetical policies, outcomes, attribute bundles, or other items.[120] For example, a survey might show the respondent pictures of current air quality versus what the air would look like if a particular regulatory policy were put into place. It might then inform the respondent that the costs of the regulation would be passed on to consumers in the form of higher prices, and ask whether the respondent would approve implementing the policy, given an expected increase of some amount in the total cost of certain goods he customarily purchases. A "yes" answer might be taken to imply that the respondent is willing to pay at least that amount for the improvement in air quality.

Although now widespread, stated-preference surveys remain controversial.[121] Fully engaging that debate is well beyond the scope of this Article. Rather, the claim I wish to make in this Section is

---

119. For overviews of this approach, see generally IAN J. BATEMAN ET AL., ECONOMIC VALUATION WITH STATED PREFERENCE TECHNIQUES: A MANUAL (2002); RICHARD T. CARSON, CONTINGENT VALUATION: A COMPREHENSIVE BIBLIOGRAPHY AND HISTORY (2011); HANDBOOK ON CONTINGENT VALUATION (Anna Alberini & James R. Kahn eds., 2006); A PRIMER ON NONMARKET VALUATION (Patricia A. Champ, Kevin J. Boyle & Thomas C. Brown eds., 2003); VALUING ENVIRONMENTAL AMENITIES USING STATED CHOICE STUDIES (Barbara J. Kanninen ed., 2007); VALUING ENVIRONMENTAL PREFERENCES: THEORY AND PRACTICE OF THE CONTINGENT VALUATION METHOD IN THE US, EU, AND DEVELOPING COUNTRIES (Ian J. Bateman & Kenneth G. Willis eds., 2001); Richard T. Carson & W. Michael Hanemann, *Contingent Valuation*, *in* 2 HANDBOOK OF ENVIRONMENTAL ECONOMICS 821 (Karl-Göran Mäler & Jeffrey R. Vincent eds., 2005); L. Venkatachalam, *The Contingent Valuation Method: A Review*, 24 ENVTL. IMPACT ASSESSMENT REV. 89 (2004).

120. A different format, less utilized at present, is an "open-ended" question that asks for the maximum the respondent is willing to pay for some good.

121. *See supra* note 119; *see also* FUJIWARA & CAMPBELL, *supra* note 11; Matthew D. Adler, *Welfare Polls: A Synthesis*, 81 N.Y.U. L. REV. 1875 (2006); Richard T. Carson, Nicholas E. Flores & Norman F. Meade, *Contingent Valuation: Controversies and Evidence*, 19 ENVTL. & RESOURCE ECON. 173 (2001); Daniel Kahneman, Ilana Ritov & David Schkade, *Economic Preferences or Attitude Expressions?: An Analysis of Dollar Responses to Public Issues*, 19 J. RISK & UNCERTAINTY 203 (1999); Daniel Kahneman & Robert Sugden, *Experienced Utility as a Standard of Policy Evaluation*, 32 ENVTL. & RESOURCE ECON. 161 (2005); Robert Sugden, *Anomalies and Stated Preference Techniques: A Framework for a Discussion of Coping Strategies*, 32 ENVTL. & RESOURCE ECON. 1 (2005).

conditional. *If* individuals have sufficiently well-behaved preferences to allow for a preference-based account of well-being, and *if* surveys are an appropriate methodology for estimating these preferences, stated-preference surveys *dominate* SWB surveys. The former are clearly better in some relevant respects, and no worse in all other.

A preference-based account of well-being requires that individuals' rankings satisfy formal rationality requirements.[122] The content of such requirements is contested, but they include—at a bare minimum—transitivity. Unless a ranking satisfies formal rationality constraints, it may not even be characterizable as a preference, and in any event, would be irrational and normatively suspect.

Some critiques of stated-preference surveys point to anomalies suggesting a violation of rationality requirements.[123] Scope insensitivity (also known as "embedding") is one such anomaly: individuals' WTP/WTA amounts do not vary appropriately with change in the scope of the good being valued. For example, a respondent's willingness to pay for a reduction in fatality risk may increase very little even as the risk reduction is doubled or quadrupled.[124] Large WTP/WTA disparities are another. The amount of money that the respondent requires in compensation for not having some good (her WTA) may be much larger than what she would pay to have it (her WTP)—suggesting that she is envisioning attribute packages as losses or gains from a reference point, arguably a kind of formal irrationality.

However, anyone proposing to use SWB surveys as evidence of *preference* utility must presuppose that such irrationalities are not widespread or entrenched (for example, that they can be overcome via sufficient debiasing or other steps to "construct"[125] well-behaved preferences). If scope insensitivities, WTP/WTA disparities, and the like really prove that most individuals are too irrational to meet the formal requirements of a preference-based account of well-being, the conclusion should be that we must reject this account entirely, thus

---

122. *See supra* text accompanying notes 35–48 (discussing the laundering of preferences).

123. *See, e.g.*, Jerry Hausman, *Contingent Valuation: From Dubious to Hopeless*, J. ECON. PERSP., Fall 2012, at 43, 46–49.

124. This is a violation of expected utility theory. *See* James K. Hammitt & John D. Graham, *Willingness To Pay for Health Protection: Inadequate Sensitivity to Probability?*, 18 J. RISK & UNCERTAINTY 33, 34–35 (1999).

125. John W. Payne, James R. Bettman & David A. Schkade, *Measuring Constructed Preferences: Towards a Building Code*, 19 J. RISK & UNCERTAINTY 243 (1999).

undermining not only stated-preference surveys, but also the PR defense of *SWB* surveys.[126]

A second, quite familiar, critique of stated-preference surveys is less radical. It asserts that behavioral ("revealed preference") data is stronger evidence of individuals' real preferences than survey data. In particular, it is claimed, individuals have strategic grounds to answer stated-preference questions insincerely. It is also suggested (a kind of nonidealization point) that respondents have no "skin in the game," no incentive to seriously consider what their preferences are, because any answer they give is costless. Stated-preference surveys are therefore (it is claimed) subject to "hypothetical bias," whereby individuals tend to articulate higher monetary values for goods than they would expend in real transactions.[127]

Undoubtedly some types of stated-preference surveys are vulnerable to strategic bias, hypothetical bias, or other slippage between respondents' genuine preferences and their answers to the survey questions. But that would also surely be true of some types of SWB surveys.[128] Nor is there any reason to believe that survey design improvements intended to mitigate these flaws are especially applicable to the SWB format.

Conversely, if the flaws just mentioned are mitigable—if properly designed surveys are a useful tool to infer preference utility—there are several strong considerations in favor of the stated-preference approach.

First, as just mentioned, preferences must satisfy formal rationality requirements. Some who advocate a preference-based account of well-being go further, arguing that the preferrer must meet various additional laundering conditions, such as being well-informed, calm and deliberative. Efforts to debias respondents, and to furnish

126. *Cf.* Dan Ariely, George Loewenstein & Drazen Prelec, *"Coherent Arbitrariness": Stable Demand Curves Without Stable Preferences*, 118 Q.J. ECON. 73, 73 (2003) (finding that market behavior may not reveal stable underlying preferences).

127. *See* CARSON, *supra* note 119, at 13 (noting that hypothetical bias and strategic misstatement are two enduring concerns about stated-preference surveys).

128. *See* Bruno S. Frey & Alois Stutzer, *The Use of Happiness Research for Public Policy*, 38 SOC. CHOICE & WELFARE 659, 670 (2012) ("When happiness indicators influence the behaviour of political actors and their policy choices, individuals have an incentive to misstate their well-being.").

DUKE LAW JOURNAL   [Vol. 62:1509

information, are a standard feature of stated-preference surveys, but not SWB surveys.[129]

Perhaps this is remediable. For example, SWB respondents could, in principle, be provided a mini-course in utility theory (so that they understand how preferences are supposed to behave, and what the ordinal or cardinal utility number attached to an attribute bundle is meant to represent). Each could be given information about her actual attribute bundle, and also asked to carefully consider whether she would prefer that bundle to various counterfactual bundles. And only *then* would respondents be asked, "How satisfied are you with your own life?" or, less elliptically, "What is your current degree of utility?" But this would represent a fairly radical change in how SWB surveys are actually conducted.

A second and more fundamental advantage of the stated-preference approach is that it is robust to preference heterogeneity—both interpersonal and intrapersonal. Howard at Time 1 ($Howard_1$) may have different preferences from Howard at Time 2 ($Howard_2$), and both of the Howards' preferences may differ from Robert's. The stated-preference approach can, in principle, accommodate this heterogeneity. $Howard_1$, $Howard_2$, and Robert can each be asked to rank a set of possible attribute bundles. If each has a different ranking, that will show up in the data.

By contrast, the SWB format presupposes some substantial degree of preference homogeneity. Even if $Howard_1$, $Howard_2$, and Robert each accurately expresses his preference utility for his current bundle, there is no way to infer how *any* of them would rank the set of possible bundles if we allow each ranking to be different.

The fundamental point is that, in the stated-preference format, *multiple* possible attribute bundles are presented to each respondent.[130] The respondent just tells us directly what his ranking is.

---

129. For one particularly intensive effort to encourage deliberation about preferences as part of a stated-preference survey, see Douglas MacMillan, Nick Hanley & Nele Lienhoop, *Contingent Valuation: Environmental Polling or Preference Engine?*, 60 ECOLOGICAL ECON. 299 (2006).

130. In practice, stated-preference practitioners often use a so-called "dichotomous choice" framework which, in effect, asks each respondent to rank only *two* attribute bundles (bundles consisting of some level of a good and some amount of income) out of a larger set. Dichotomous-choice questions, besides having desirable properties as regards strategic bias, are less cognitively demanding than questions asking the respondent for fuller ranking information. Answers to a survey posing each respondent a dichotomous-choice question over some pair of bundles, *together with assumptions about the homogeneity of respondents' preferences*, can be used to infer the common preferences over the entire set. But it is also quite possible to ask each

In the SWB format, each member of a group is asked to articulate his preference utility for *one* bundle (his current one). Information about counterfactuals—namely, whether an individual would prefer or disprefer his current bundle to a different one—is *not* elicited. Without such information, we can *only* infer whether someone would prefer or disprefer his current bundle to a different one *if* we assume that his preferences are to some degree homogeneous with the holder of that other bundle.[131]

Third, and equally importantly, stated-preference surveys do not ask the respondent to translate her ranking of bundles into a utility scale and to express her ranking by telling us her utility. Instead, she is just asked for the ranking. The whole problem of scale heterogeneity, a deep difficulty for SWB surveys (as evidence of preference utility), is a non-issue for the stated-preference format.

This is certainly true if we allow for interrespondent preference heterogeneity. And it is true even if we *do* combine the stated-preference format with assumptions about preference homogeneity so as to increase the inferential power of this approach.

For example, imagine that we try to infer how each in a group of respondents would rank a fairly large set of attribute bundles. It seems overwhelming to present the whole set to each, and so we assume that all have the same preferences. With this assumption in hand, we could administer a stated-preference survey, asking each respondent to rank a particular subset of the larger set. With this information in hand, together with formal features of preferences (transitivity) and the homogeneity assumption, we can reach a conclusion about how each in the group ranks the larger set. Nothing in this conclusion depends upon a *further* assumption that each in the group expresses these common preferences via a common utility scale.

respondent to rank three or more bundles, or to pick the most preferred of three or more bundles; and stated-preference surveys sometimes *do* pose such questions, as in the so-called "choice experiment" or "contingent ranking" format.

131.   Some critics of using stated-preference surveys to estimate WTP/WTA values have argued that these surveys impose a cognitive burden by requiring respondents to rank multiple bundles—and have suggested that such burden is reduced by the SWB format. *See, e.g.*, Frey et al., *supra* note 11, at 148. But this cognitive-burden defense of the SWB format is persuasive only if SWB surveys are taken as evidence of something about respondents *other than their preferences*. The "burden" of ranking multiple alternatives is an inextricable part of having a preference.

Fourth, stated-preference surveys are less susceptible to the miscommunication difficulties that affect SWB surveys. It is hard to misunderstand the question, "Which would you choose?" or "Which do you prefer?" By contrast, the person asked "How satisfied are you with your life?" or "What is your current utility?" (let alone, "How happy are you?") might well misunderstand the question as requesting a measure of affect or good feeling—experience utility—rather than a measure of preference realization.

III. SUBJECTIVE WELL-BEING SURVEYS AS EVIDENCE OF EXPERIENCE UTILITY

The PR defense sees SWB surveys as evidencing preference utility, a measure of the extent to which individuals' preferences are realized (with individuals permitted to have an intrinsic preference for items that are at least partly nonexperiential, such as consumption, health, liberty, accomplishment, knowledge, and so forth). By contrast, the EQ defense argues that SWB surveys provide useful information about the quality of individuals' mental states. This defense is, in a way, much more straightforward. Although individuals are not infallible about the content of their mental states, surely each individual is generally more epistemically reliable about what she thinks and feels than about the occurrence of what she wants. Moreover, it seems straightforward to design survey questions focusing on experiential quality: for example, "How happy are you?"

Indeed, many scholars in the SWB literature offer (what appears to be) some version of the EQ defense. The most prominent example is Daniel Kahneman, who argues that information about experience utility should play a substantial role in structuring governmental choices. Kahneman and his collaborators (for short, "Kahneman") have pioneered a novel framework ("objective happiness") for measuring experience utility, and have empirically implemented this framework in several large-scale studies.[132] But Kahneman is hardly alone in seeing SWB surveys as evidencing experience utility. There are numerous other SWB researchers who present—or at least seem to present—the EQ defense.

That defense, once more, can take a strong or weak form. In the strong form, the EQ defense of SWB surveys endorses experientialism about well-being. A leading example of this approach

---

132. *See infra* Part III.B.

is the best-selling book, *Happiness: Lessons from a New Science*,[133] written by Richard Layard, a prominent SWB researcher. Section A responds to Layard's arguments and, more generally, criticizes the strong EQ defense.

The weak EQ defense is more promising. It refrains from endorsing experientialism about well-being. Experientialism about well-being is, at a minimum, normatively controversial. The weak EQ defense of SWB surveys declines to take sides in that thorny debate. It claims only that good experiences are one *aspect* of well-being—a claim which seems very hard to deny.

More problematic is the assertion that policymakers should take account of the experiential impact of governmental policies via SWB surveys. In Section B, I illustrate the difficulties with the weak EQ defense of SWB surveys via a close analysis of Kahneman's "objective happiness" framework. Kahneman's work is by far the most systematic attempt, to date, to develop a policy-relevant measure of mental states. And Kahneman now acknowledges (or at least is willing to entertain) that well-being has nonmental aspects. The "objective happiness" framework should thus be understood as a concrete elaboration of the weak EQ defense: one that sees SWB surveys as evidence of experiential quality, and experiential quality as one important determinant (among others) of well-being.

Key objections to the "objective happiness" framework are its implausible presuppositions regarding temporal separability and, even more fundamentally, the separability of the hedonic from the nonhedonic. Relatedly, the framework offers no guidance in how policymakers should integrate information about hedonic utility with nonhedonic information. Finally, although the theoretical elaboration of "objective happiness" uses an "observer" to make time-tradeoff judgments so as to cardinalize hedonic utilities, the "observer" is not to be found in Kahneman's empirical studies, and even the theoretical elaboration fails to allow for heterogeneity in observer judgments/preferences.

Section C sketches a different and arguably more promising approach to integrating information about experiential quality into policy choice: an approach that includes happiness as *one* of the entries in individuals' preference utility functions, and that employs

---

133.   LAYARD, *supra* note 10.

revealed or stated-preference studies to estimate the extent to which preference utility depends upon happiness.

A. Experientialism about Well-Being: The "Strong" Experience-Quality Defense of Subjective Well-Being Surveys

Experientialism about well-being has a distinguished intellectual history. Bentham, of course, was an experientialist, and so (in a different way) were John Stuart Mill and Henry Sidgwick.[134] But, more recently, the view fell into philosophical disfavor. The most influential twentieth-century work of moral philosophy, John Rawls's *A Theory of Justice*,[135] adopts a preference-based view of welfare, rather than equating well-being with pain and pleasure or, more generally, good experiences.[136] The other leading contemporary philosophical works on well-being (here, I am thinking of the work of Richard Brandt,[137] James Griffin,[138] John Finnis,[139] Martha Nussbaum,[140] and Wayne Sumner[141]) are also nonexperientialist.

Indeed, for a time, Robert Nozick's discussion of the "experience machine" was generally seen by philosophers as a decisive refutation of experientialism.[142] Over the last decade or so, *some* scholars have pushed back against this conventional wisdom.[143] Still, it remains the case that experientialism about welfare is a philosophically controversial position. As already explained in Part I, two widely accepted classes of well-being accounts—preferentialism (with the special exception of a preferentialist view that embraces an

---

134. SUMNER, *supra* note 26, at 83–92.

135. JOHN RAWLS, A THEORY OF JUSTICE (rev. ed. 1999).

136. *See id.* at 358–80.

137. *See generally* RICHARD B. BRANDT, A THEORY OF THE GOOD AND THE RIGHT (1979).

138. *See generally* JAMES GRIFFIN, WELL-BEING: ITS MEANING, MEASUREMENT, AND MORAL IMPORTANCE (1986).

139. *See generally* FINNIS, *supra* note 42.

140. *See generally* NUSSBAUM, *supra* note 31.

141. *See generally* SUMNER, *supra* note 26.

142. *See, e.g.*, Sharon Hewitt, *What Do Our Intuitions About the Experience Machine Really Tell Us About Hedonism*?, 151 PHIL. STUD. 331, 332 (2010); Matthew Silverstein, *In Defense of Happiness: A Response to the Experience Machine*, 26 SOC. THEORY & PRAC. 279, 281–82 (2000).

143. *See, e.g.*, ROGER CRISP, REASONS AND THE GOOD (2006); FRED FELDMAN, PLEASURE AND THE GOOD LIFE: CONCERNING THE NATURE, VARIETIES, AND PLAUSIBILITY OF HEDONISM (2004); FRED FELDMAN, WHAT IS THIS THING CALLED HAPPINESS? (2010); Mark Bernstein, *Well-Being*, 35 AM. PHIL. Q. 39 (1998); Torbjörn Tännsjö, *Narrow Hedonism*, 8 J. HAPPINESS STUD. 79 (2007).

experientialist conception of "self-interested"[144]) and the objective-good approach—reject the experientialism requirement. Accounts within these classes reject the proposition that the *only* way to directly change someone's well-being is by changing the content of her mental states (what she thinks, feels, remembers, and so forth).

In *Happiness*, however, Layard adopts a Benthamite view of well-being and of the proper goals for social policy. He writes, "I believe that Bentham's idea"—that "all laws and all actions should aim at producing the greatest possible happiness"—"was right and that we should fearlessly adopt it."[145]

Happiness almost completely ignores the contemporary philosophical debates about well-being. Perhaps that is not surprising, given academic specialization (Layard is an economist) and the book's aim to reach a popular audience. Still, one can ask: has Layard advanced or at least sketched plausible arguments for why governmental policy should be *solely* focused on promoting individual happiness, rather than nonexperiential features of individual lives (except as an instrumental means to more happiness)?

Layard argues as follows:

> Why should we take the greatest happiness as the goal for our society? Why not some other goal—or indeed many? What about health, autonomy, accomplishment or freedom? The problem with many goals is that they often conflict, and then we have to balance one against the other. So we naturally look for one ultimate goal that enables us to judge other goals by how they contribute to it.

> Happiness is that ultimate goal because, unlike all other goals, it is self-evidently good. If we are asked why happiness matters, we can give no further, external reason. It just obviously does matter. As the American Declaration of Independence says, it is a "self-evident" objective.

> By contrast, if I ask you why you want people to be healthier, you can probably think of reasons why—people should not be in pain, they should be able to enjoy life and so on. Similarly, if I ask you about autonomy you will point out that people feel better if they can

---

144. *See supra* text accompanying note 48; *infra* notes 163–175 and accompanying text.

145. LAYARD, *supra* note 10, at 111–12. Layard's book was originally published in 2005. Citations are to the revised and updated edition, published in 2011, which added a new part but did not change the main text. *See id*. at xiii. The new part adds little to the original argument for happiness maximization, except a slight addendum to the argument about the uniquely self-evident intrinsic value of happiness. *See infra* note 154.

control their own lives. Likewise, freedom is good because slavery, prison and the secret police lead to nothing but misery.

So goods like health, autonomy and freedom are "instrumental" goods—we can give further, more ultimate reasons for valuing them. . . .

To help us promote the greatest happiness, we obviously need to understand what conditions affect people's happiness, and by how much. This is now becoming possible on an empirical basis. . . . But unless we can justify our goals by how people feel, there is a real danger of paternalism. We ought never to say: this is good for you, even though it will never make you or others feel better. On the contrary, if we want to measure the quality of life, it must be based on how people feel.[146]

After addressing objections to happiness-maximization, Layard rearticulates his position in this way: "So we come back to the central idea of a humane ethic that values what people want for themselves, for their children and for their fellow citizens. And that is their happiness."[147]

In these passages, one can tease out three separate lines of defense of experientialism about well-being. None are particularly persuasive.

(1) *Monism*: "The problem with many goals is that they often conflict."[148] A view that equates well-being and happiness is unidimensional or "monistic." By contrast, if we allow well-being to depend upon both happiness *and* nonexperiential items such as health, autonomy, accomplishment or freedom, it becomes multidimensional. On such an account, how is government supposed to choose a course of action when the dimensions conflict?

What this argument overlooks is that there are very well-developed techniques in economics for assigning a given individual a single utility number as a function of her attainment on multiple dimensions.[149] Conversely, experientialism is hardly a sure recipe for avoiding the complexities of multidimensionality. On the most plausible accounts, experiential quality is *itself* multidimensional: a matter of good perceptions, cognitions, and memories, not just pains

---

146. *Id.* at 112–13 (footnote omitted).

147. *Id.* at 124.

148. *Id.* at 112.

149. *See generally* RALPH L. KEENEY & HOWARD RAIFFA, DECISIONS WITH MULTIPLE OBJECTIVES: PREFERENCES AND VALUE TRADEOFFS (1993).

and pleasures. Many SWB researchers see subjective well-being (good mental states) as a composite of positive and negative affect *and* feelings of satisfaction.[150] Even the narrower, Benthamite view may not really be monistic. As discussed at greater length below, painfulness and pleasantness may turn out to be *separate* dimensions.[151]

(2) *Only Happiness Has Self-Evident Intrinsic Value*: "Happiness is that ultimate goal because, unlike all other goals, it is self-evidently good. . . . [G]oods like health, autonomy and freedom are 'instrumental' goods—we can give further, more ultimate reasons for valuing them."[152] Layard's idea of "self-evidence" invokes a standard feature of normative reasoning, namely that some normative propositions might be supported by direct appeal to intuitions.[153] Layard is surely correct that many of his (and this Article's) readers *will* strongly intuit that happiness has intrinsic welfare value: happiness, as such, increases well-being. But many readers will *also* intuit that physical health, autonomy, knowledge, relationships, or accomplishment have intrinsic welfare value. This relates to pluralism/monism: there is nothing at all confused in having the intuition that multiple types of individual attributes, both happiness and other goods, are intrinsic welfare components.

In characterizing happiness as a "self-evident" good, Layard not only appeals to intuitions in favor of happiness, but also emphasizes the strength of such intuitions. Assume, arguendo, that intuitions concerning the welfare relevance of health (for example) are generally weaker than intuitions concerning the welfare relevance of happiness. It would be a fallacy to conclude that only happiness, not health, has intrinsic welfare value. If I intuit that proposition P^* is true, then (ceteris paribus) my normative views should be consistent with P^*, even if I hold the intuition with less than complete certainty, and even if there is some other proposition P (logically consistent with P^*) that I intuit more strongly.

In principle, Layard *might* try to undercut intuitions concerning the welfare relevance of nonexperiential items in a different way, by

---

150. For a fuller discussion of the multidimensionality of good experiences, see *infra* notes 197–200 and accompanying text.

151. *See infra* Part III.B.1.

152. LAYARD, *supra* note 10, at 113.

153. *See* NORMAN DANIELS, JUSTICE AND JUSTIFICATION: REFLECTIVE EQUILIBRIUM IN THEORY AND PRACTICE (1996).

arguing that (a) such intuitions do not survive idealization (with full information, and thinking rationally, no one intuits that health, etc., has intrinsic welfare value), or that (b) such intuitions are much less widely shared than the intuition regarding the value of happiness. The first tack seems a nonstarter—undercut by the many serious thinkers who have embraced nonexperientialist accounts. The second tack is empirically speculative. Layard doesn't present (or attempt to present) evidence to support it.[154] And evolutionary considerations actually suggest that intuitions supporting the intrinsic significance of at least some nonexperiential items *are* likely to be quite widely shared.[155]

Layard's key argument against the intrinsic value of health, autonomy, and freedom is yet another one: namely, that we can give *instrumental* reasons for promoting these items. This argument is fallacious. Something can be *both* intrinsically *and* instrumentally valuable. In particular, imagine that there are two intrinsic components of welfare: being happy and being healthy. Then we *can* reason instrumentally about health (asking about the effect of health-promotion policies on happiness), but we can also point to the health-promotion effect of a policy as an intrinsic reason to support it. Symmetrically, we can reason instrumentally about *happiness* by asking about the effect of happiness-promotion policies on *health*. The health benefits of affects and cognitions are in fact a large topic of research among public health researchers![156]

Thus, the premise that nonexperiential items possess instrumental value, as causal precursors to happiness, does *not* imply the conclusion that they lack intrinsic value.

(3) *Paternalism*: "[U]nless we can justify our goals by how people feel, there is a real danger of paternalism."[157] Government acts paternalistically toward some individual when it restricts her choices *for her own good*—when it limits what she can do, not because of third-party effects, but because it believes that she will fail to promote her own well-being.

---

154. In the new part to his book, added in the revised and updated edition, Layard claims—without empirical support—that happiness "is the only good which would be *generally accepted* as an end in itself." LAYARD, *supra* note 10, at 240.

155. *See* ELLIOTT SOBER & DAVID SLOAN WILSON, UNTO OTHERS: THE EVOLUTION AND PSYCHOLOGY OF UNSELFISH BEHAVIOR 296–327 (1998).

156. *See* Diener & Seligman, *supra* note 10, at 13–15.

157. LAYARD, *supra* note 10, at 113.

Far from being antipaternalistic, experientialism has the real potential for paternalism.[158] If citizens self-interestedly prefer nonexperiential items, a government with a mandate to promote good experiences has a justification for restricting or altering their choices. For example, if mental health interventions are especially conducive to happiness, but some individuals have a substantial intrinsic desire to be physically healthy, an experientialist government might try to tilt those individuals' health expenditures away from their preferred mix as between physical and mental health. It is *preference* views—drawing an equivalence between what someone wants (under stipulated conditions) and what is good for that person—that have the conceptual resources to resist paternalism.[159] Conversely, the possible disjunction between the course of action that maximizes the realization of someone's self-interested preferences, and the course of action that maximizes the quality of her feelings or experiences, renders an experiential account potentially paternalist.

Layard obscures this potential via his assertion that individuals self-interestedly prefer their own happiness. "So we come back to the central idea of a humane ethic that values what people want for themselves, for their children and for their fellow citizens. And that is their happiness."[160] But this is an empirical assertion—one that Layard does not back up with evidence, and one that the studies reviewed in Part I.B. do not clearly support.

Of course, the sheer fact that a well-being account is paternalistic does not provide a decisive reason to reject it. Intuitively, some individuals *do* fail to promote their own well-being. But SWB scholars such as Layard should not inflate the virtues of experientialism by obscuring where it stands on paternalism.

I have parsed Layard's brief for experientialism and found it wanting. But better arguments for his position might be available. John Bronsteen, Christopher Buccafusco, and Jonathan Masur have recently offered a lengthy defense of experientially focused policy

---

158. Robert Sugden, *Capability, Happiness, and Opportunity*, *in* CAPABILITIES AND HAPPINESS 299, 300 (Luigino Bruni, Flavio Comim & Maurizio Pugno eds., 2008).

159. To be sure, preference views become increasingly paternalistic as they impose increasingly stringent rationality and informational conditions on preferences. The preference component of any such view pushes against paternalism, whereas the idealization component pushes toward it.

160. LAYARD, *supra* note 10, at 124.

analysis.[161] And, as mentioned, some contemporary academic philosophers have tried to mount a case for (or at least deflate some of the standard objections to) experientialism.[162]

To my mind, the strongest reason to endorse an experientialism requirement is a reason that Layard does not mention: the "good-for" aspect of welfare.[163] Well-being is *subject-relative* goodness. Some occurrence enhances Sonya's well-being only if it is good *for Sonya*, rather than merely being good in an impersonal sense, or good for someone else. Call this the "non-remoteness" constraint on well-being: an occurrence has a direct[164] impact on someone's welfare only if it is not too "remote" from her.

At first blush, the non-remoteness constraint presents a plausible case for experientialism. As Shelly Kagan explains:

> Suppose I meet a stranger on a train. He tells me his story, and I form the desire that he succeed in his projects. We then part, and I never hear of him or even think of him again. If he does in fact succeed, then my desire has been satisfied. According to the desire theory, then, this makes me better off. But this is intuitively an absurd claim. Obviously my level of well-being is not affected *at all* by the success of the stranger. . . .

> This suggests that the unrestricted desire theory [of well-being] is hopelessly broad. A theory of well-being must explain which facts constitute my being better off. So they must be facts about *me*. Since my desires can range over facts that have nothing whatsoever to do with me, the satisfaction of such desires cannot constitute my well-being. . . .

> From this perspective, the position of mental statism no longer seems so arbitrary. At least it seems to keep the content of well-being within reasonable bounds, for facts about my mental states are certainly facts about me. In contrast, it is far from clear whether anything external to my mind . . . can count as—in the relevant

---

161. *See, e.g.*, Bronsteen, Buccafusco and Masur, *Welfare as Happiness*, *supra* note 10; Bronsteen, Buccafusco and Masur, *Well-Being Analysis*, *supra* note 10.

162. *See supra* note 143.

163. *See* SUMNER, *supra* note 26; Bernstein, *supra* note 143, at 45–46; Shelly Kagan, *The Limits of Well-Being*, SOC. PHIL. & POL'Y, June 1992, at 169, 171–72.

164. The word "direct" is needed because remote occurrences might have an instrumental, causal, impact on someone's well-being.

sense—facts about me. If not, then the limits of well-being must be drawn at the limits of our minds.[165]

The "stranger" case described by Kagan derives originally from philosopher Derek Parfit.[166] Eric Posner and I provided a similar case in our book on cost-benefit analysis, involving a person named Sheila and an endangered species, the Sri Lankan squirrel.

> One outcome involves . . . the continued existence of the Sri Lankan squirrel . . . ; the other is the extinction of this species. Sheila has never traveled to Sri Lanka, and never intends to, nor is she an environmentalist who's made species preservation her life's work, but she still (slightly, say) prefers the first outcome because she believes that morality includes environmental values disfavoring the disappearance of species. It seems odd to say that the nonextinction of the squirrel makes Sheila better off.[167]

The "stranger" and Sri Lankan squirrel cases underscore the non-remoteness constraint on well-being. More specifically, within the confines of a preference-based account of welfare, they show the need to restrict the category of welfare-relevant preferences. The cases illustrate that someone might have a fully laundered preference (a preference that is well-informed, rational, intelligible, and so forth) for outcome x over y, and still not be better off in x than y. In order to survive the remoteness objection, the preference-based account must say that person i is better off in x than y if and only if i has a *self-interested*, laundered preference for x over y. The reasons thus to restrict the category of welfare-relevant preferences include not just our intuitive reactions to particular hypothetical cases, but also deeper considerations regarding the possibility of self-sacrifice and of moral reasoning.[168]

Kagan concludes, however, that the non-remoteness constraint does not, on balance, argue for experientialism. I agree with Kagan. In particular, the literature on preferentialism has offered at least four possible general solutions to the problem of differentiating self-interested from non-self-interested preferences.[169] (1) *Experientialism*: An individual's preference is self-interested if and only if the

---

165.   Kagan, *supra* note 163, at 171–72.
166.   *See* DEREK PARFIT, REASONS AND PERSONS 494 (1987).
167.   ADLER & POSNER, *supra* note 32, at 34.
168.   *See id.* at 34–35; ADLER, *supra* note 27, at 174–78.
169.   *See* ADLER, *supra* note 27, at 178–81.

fundamental arguments[170] for the preference are the individual's mental states. (2) *Experientialism plus physicalism*: An individual's preference is self-interested if and only if the fundamental arguments for the preference are the individual's mental states or physical states. (3) *Existence-entailment*: An individual's preference is self-interested if and only if the realization of the preference entails the individual's existence. (4) *Sympathy*: An individual's preference is self-interested if and only if based in self-sympathy, an attitude of care and concern for herself.

The experientialist conception of "self-interested" strikes me as too narrow. Surely occurrences within a person's physical body are not "remote" from her. As Kagan explains, even if we insist that a preference is self-interested only if its fundamental arguments are the person's nonrelational attributes, that would lead us to conception (2) in the previous paragraph, not (1).[171] We get from a nonrelationality requirement to experientialism only by adopting a controversial view of personal identity—namely, that a particular human person is *just* a bundle of psychological properties, rather than a particular human being who has a brain *and* a body.[172] Moreover, I would go wider than experientialism plus physicalism. Given that practical rationality, theoretical rationality, and affiliation are three central capacities of human persons—the capacities to make choices in the pursuit of goals, to acquire knowledge, and to form relationships with others—it seems to me that preferences for success in the exercise of those capacities can count as self-interested even if outside the scope of conceptions (1) and (2). Such preferences are not merely "intelligible" (as I discussed in Part I),[173] but intelligible *as a matter of self-interest*. A second case that Posner and I discussed in our book, the case of the deceived scholar, illustrates this point.[174]

---

170. *See supra* notes 31–34 and accompanying text.

171. Kagan, *supra* note 163, at 180–89.

172. *See generally* DAVID DEGRAZIA, HUMAN IDENTITY AND BIOETHICS (2005); ERIC T. OLSON, THE HUMAN ANIMAL: PERSONAL IDENTITY WITHOUT PSYCHOLOGY (1997).

173. *See supra* notes 34–47 and accompanying text.

174. ADLER & POSNER, *supra* note 32, at 30. Let me modify the case somewhat to highlight the connection with goals, knowledge, and relationships. Imagine that David, an academic scientist, works hard at his research. He hopes to make a scientific discovery and to be respected for doing so by his colleagues. In considering possible outcomes, David says that he prefers *x* (an outcome in which he actually discovers some new and significant scientific fact), as opposed to *y* (an outcome in which he falsely believes to have done so, abetted in this belief by colleagues who want to spare his feelings but pity David behind his back). Even though David's mental states are identical in these possible outcomes, there are various important respects in

Bronsteen, Buccafusco, and Masur argue that I have "directly contradict[ed]" myself in denying that the squirrel's survival benefits Sheila, and yet rejecting the experientialist conception of self-interest, conception (1).[175] Their characterization, if accurate, would also hold true of anyone else who adopts conception (2), (3), (4), or for that matter any view which analyzes self-interested preferences other than as preferences for experiences. But, in fact, there is no contradiction here: just a position intermediate between the narrow insistence that what is good for me must occur within my head, and an overbroad willingness to include the realization of any preference (including Sheila's, or a preference for the stranger's success in Parfit's case) as improving *my* welfare.

The brief discussion of the last several paragraphs will hardly satisfy the proponent of experientialism. The non-remoteness problem is very difficult, and the experientialist solution to that problem certainly *does* deserve close and serious consideration. Although I believe that this argument for experientialism fails—as do others advanced by contemporary philosophical experientialists, and by Bronsteen, Buccafusco, and Masur—I lack space to pursue the analysis here. Notwithstanding Nozick's "experience machine," experientialism should not be rejected out of hand.

---

which his preference seems self-interested whereas Sheila's does not: in *y* he has failed at his career goal, he knows less (and has made no contribution to human knowledge), and a thread of deceit runs through his relationships with his colleagues.

Note that both conceptions (3) and (4) would count David's preference for *x* as self-interested. A preference that *I* contribute to finding a cure for cancer—by contrast with a preference that a cure for cancer be discovered—is existence-entailing, because all outcomes in which I do not exist are ranked equal by the first but not the second preference. David's preference is that *he* make a scientific breakthrough and that *he* be respected by his colleagues as a result. Moreover, someone caring about David (whether David himself, or someone else) would be motivated to pursue *x*. For example, if David has a sympathetic friend who knows of David's capacity for self-deception, the propensities of David's colleagues, and so forth, and believes that David's research is heading down a false path—and thus taking David in the direction of *y* not *x*—the friend might try to get David to redirect his research.

175.   Bronsteen, Buccafusco and Masur, *Welfare as Happiness*, *supra* note 10, at 1621. They also point out that an account of self-interest which categorizes as "self-interested" the deceived scholar's preference for genuine academic success (as opposed to the mere belief thereof) also thus categorizes the preference for success of a "driven scholar" willing to sacrifice his happiness and family life. *Id.* at 1625–26. This is true. But it is a further question whether this latter preference survives full information and rational reflection on the driven scholar's part, and yet a further question whether it would be widely shared (a point relevant to interpersonal comparisons, *see* ADLER, *supra* note 27, at 185–225). Note, here, that the critic of the experientialism requirement has a pretty easy argumentative burden. What she needs to show is that there are *some* cases in which someone's well-being is directly improved by nonexperiential changes, *not* that such changes always override experiential losses.

Reciprocally, however, SWB scholars tempted to equate well-being and subjective well-being should recognize that they are entering a hornet's nest of disputation, and should be prepared to do serious normative battle in defense of experientialism. Lacking the appetite or ammunition for such battle, the SWB scholar might instead claim that good experiences are at least one aspect of well-being, whether or not the sole determinant.

That claim is hard to dispute. But does it justify the use of SWB surveys as a basis for policy choice? It is to that question that we now turn.

B. *"Objective Happiness" and the "Weak" Experience-Quality Defense*

This Section discusses Kahneman's "objective happiness" approach.[176] I remind the reader that "Kahneman" is shorthand for Kahneman *and* the distinguished collaborators with whom he has co-authored the theoretical and empirical articles developing this framework.

Kahneman disavows (or at least does not commit himself to) an experientialist conception of well-being. He writes:

> Defining happiness by the temporal distribution of experienced affect appears very narrow, and so it is. The concept of objective happiness is not intended to stand on its own and is proposed only as a necessary element of a theory of human well-being. A comprehensive account of well-being inevitably brings in philosophical considerations and a moral conception of "the good life," which are not easily reduced to experienced utility. However, good mood and enjoyment of life are not incompatible with other

---

176. For the theoretical elaboration of the objective-happiness framework, see Daniel Kahneman, Peter P. Wakker & Rakesh Sarin, *Back to Bentham? Explorations of Experienced Utility*, 112 Q.J. ECON. 375 (1997); Daniel Kahneman, *Experienced Utility and Objective Happiness: A Moment-Based Approach*, *in* CHOICES, VALUES, AND FRAMES, *supra* note 70, at 673 [hereinafter Kahneman, *Experienced Utility*]; Daniel Kahneman & Jacob Riis, *Living, and Thinking About It: Two Perspectives on Life*, *in* THE SCIENCE OF WELL-BEING, *supra* note 1, at 285; and Daniel Kahneman, *Objective Happiness*, *in* WELL-BEING: THE FOUNDATIONS OF HEDONIC PSYCHOLOGY, *supra* note 1, at 3, 3–12 [hereinafter Kahneman, *Objective Happiness*]. For empirical implementation, see generally MEASURING THE SUBJECTIVE WELL-BEING OF NATIONS, *supra* note 4; Kahneman et al., *supra* note 7; and Alan B. Krueger et al., *Time Use and Subjective Well-Being in France and the U.S.*, 93 SOC. INDICATORS RES. 7 (2009). For further discussions of the framework, see Dolan & Kahneman, *supra* note 56; Daniel Kahneman & Alan B. Krueger, *Developments in the Measurement of Subjective Well-Being*, J. ECON. PERSP., Winter 2006, at 3; Kahneman et al., *supra* note 10; Kahneman & Sugden, *supra* note 121.

> psychological criteria of well-being that have been proposed, such as the maintenance of personal goals, social involvement, intense absorption in activities, and a sense that life is meaningful. Clearly, a life that is meaningful, satisfying, and cheerful should rank higher on the scale of well-being than a life that is equally meaningful and satisfying but sad or tense. Objective happiness is only one constituent of the quality of human life, but it is a significant one.[177]

Kahneman, here and elsewhere, tries to bolster the normative appeal of "objective happiness" by stressing that it is "only one constituent of the quality of human life."[178] The framework thus merits close critical attention, not only in its own right, but as a concrete elaboration of the weak EQ defense of SWB surveys.

The framework offers a methodology for measuring the *hedonic* or *affective* aspect of individual experience: whether a mental state is affectively positive or negative, that is, painful or pleasurable. (Kahneman often uses "hedonic," "affective," and "pain"/"pleasure" interchangeably, and I will follow his usage here.) An experience is hedonically/affectively positive, according to Kahneman, if it feels good and if the individual wants the experience to continue. An experience is hedonically/affectively negative if it feels bad and the individual wants it to stop.

The formal model underlying "objective happiness" presumes that there is a neutral hedonic level, and that a given individual (the subject), at a given moment, can meaningfully characterize her current, momentary, mental state as affectively positive, negative, or neutral—as hedonically better than, worse than, or equally good as the neutral level. Moreover, she can order possible momentary experiences *within* each of the two affective domains. Given two affectively positive experiences, the first can be ranked as more, less, or equally positive as the second—various momentary pleasures are not just pleasures simpliciter (better than neutral), but more or less pleasurable. Similarly, the subject can determine whether one hedonically negative momentary experience is more or less negative than another.

---

177.   Kahneman, *Experienced Utility*, *supra* note 176, at 683 (citations omitted). For other passages in which Kahneman declines to commit himself to an experientialist account of well-being, see Dolan & Kahneman, *supra* note 56, at 229–30; Kahneman et al., *supra* note 176, at 377 & n.3, Kahneman & Sugden, *supra* note 121, at 178 n.11; Kahneman & Riis, *supra* note 176, at 288–89.

178.   Kahneman, *Experienced Utility*, *supra* note 176, at 683.

These intra-domain hedonic rankings of momentary experiences can be represented (Kahneman assumes) by numbers—by momentary hedonic utilities—which, thus far, are merely ordinal. If the subject assigns -3 to momentary experience *A*, and -6 to momentary experience *B*, those numbers mean that both are worse than neutral, and that *B* is more negative than *A*; the numbers -3 and -8 would equally well capture this hedonic information. In order to cardinalize momentary utilities, Kahneman introduces an "observer."[179] The observer ranks temporally extended episodes ("profiles"), each consisting of a series of affective states (positive, negative, or neutral) experienced for some length of time. The observer's ranking of the profiles is consistent with axioms of time neutrality, monotonicity, and separability. Time neutrality says that adding neutral time before or after a profile does not change where the profile is located in the observer's ranking. Monotonicity says that replacing one or more positive momentary experiences within a given profile with more intensively positive momentary experiences must improve how the profile is ranked by the observer, and symmetrically for negative experiences. And separability says that the ranking of two profiles does not depend on what is experienced during moments when experiential quality is the same.[180]

If the observer's ranking satisfies these axioms, then there is a *cardinal* measure of momentary hedonic value—call it the $v(.)$ function—such that the observer ranks one profile over another if and only if the sum of the duration-weighted $v(.)$ values for the first profile is greater than the sum of the duration-weighted $v(.)$ values for the second. For example, let profile *P* consist of experiences *A*, *B*, and *C*, each for two units of time, and let *P** consist of experience *D* for five units. Then if the observer prefers *P* to *P**, it will be the case that $2v(A) + 2v(B) + 2v(C) > 5v(D)$.

The $v(.)$ function is cardinal in the sense that its values are added together to yield the overall value of a profile. The $v(.)$ function,

---

179.   *See* Kahneman et al., *supra* note 176, at 388–89, 91; Kahneman, *Experienced Utility*, *supra* note 176, at 680–81; Kahneman, *Objective Happiness*, *supra* note 176, at 5–6.

180.   It might seem absurd to compare profiles with short temporal duration. For example, if *P* consists of an hour of certain experiences, and *P** an hour of different experiences, is the observer meant to contemplate a life that endures for only an hour? Or (almost as absurdly) a life that is entirely neutral except for *P*, as compared to a life entirely neutral except for *P**? Given separability, this problem can be avoided. The observer can rank *P* and *P** by contemplating some arbitrary profile *M* followed by *P*, as compared to the very same profile *M* followed by *P**.

strictly speaking, is not unique, but unique up to a positive ratio transformation. Thus the summation or averaging of momentary $v(.)$ values—and for that matter, the summation or averaging of the overall values of profiles—is a meaningful operation.[181] Because the $v(.)$ function is cardinal, it can serve as the theoretical basis for the policy metrics that Kahneman recommends—gross national happiness and other such metrics that add or average measures of experiential quality across moments and persons.

To see how information about the observer's ranking allows us to arrive at a cardinal $v(.)$ function, return to the case in which the subject experiences A and B as both negative, and A as less negative than B. Although this hedonic data—without more—is not sufficient to determine whether the numbers assigned to the two experiences are -3 and -6, or -3 and -8, imagine that we now learn that the observer is indifferent between a profile with T hours of A experience and a profile with $T/2$ hours of B experience. This information constrains the $v(.)$ value of A to be a negative number which is half the $v(.)$ value of B. Not only must it be the case that $v(B) < v(A) < 0$—which follows immediately from the fact that both feel painful to the subject, and B feels more painful. In order to represent the *observer*'s indifference between the two profiles just mentioned, it must be the case that $v(A)T = v(B)T/2$, i.e., $v(A)/v(B) = 1/2$. This fact about the observer's ranking rules out the pair -3, -8.

To reiterate, the $v(.)$ function is identified as *that* cardinal function which both respects the hedonic intensity of momentary affective experiences, as registered by the subject, *and* additively represents the observer's ranking of all possible temporally extended episodes. The observer's ranking of temporally extended episodes, *as well as* the subject's felt experience of momentary pains and pleasures, plays a critical role—in Kahneman's theoretical apparatus—in arriving at $v(.)$.

---

181. Assume that $v(.)$ is such that its duration-weighted values represent the observer's ranking of profiles. Let $v^*(.)$ be defined as follows: for every momentary experience A, $v^*(A) = av(A)$, where a is a positive constant. Note now that duration-weighted $v^*(.)$ values equally well represent the observer's ranking of profiles. Why? Let $(A, B, …)$ be a series of momentary experiences and $(A', B', ….)$ a different series, with each experience A enduring for time $t(A)$. Then whenever $v(A)t(A) + v(B)t(B) + … \geq v(A')t(A') + v(B')t(B') + …$, it will also be true that $v^*(A)t(A) + v^*(B)t(B) + … \geq v^*(A')t(A') + v^*(B')t(B') + ….$ Kahneman goes further, showing that $v^*(.)$ represents the observer's ranking not just *if*, but *only if*, it is a positive ratio transformation of $v(.)$. Kahneman et al., *supra* note 176, at 398–402.

The "objective happiness" framework, although an important advance, has substantial limitations as a method for integrating information about experiential quality into policy analysis—limitations either in the theoretical statement of the framework or in its empirical implementation by Kahneman. Those limitations, I believe, are the following.

1. *Is the Hedonic Aspect of Experience Unidimensional or Bidimensional?* Kahneman sees hedonic value as unidimensional. Pleasure and pain are twin poles of a single spectrum bifurcated at hedonic neutrality, just as positive and negative numbers correspond to two segments of a single line running through the number 0.

This assumption of the unidimensionality of affective experience is undermined by substantial psychological evidence for a *two*-dimensional (or, more generally, higher dimensional) affective space.[182] Such evidence includes (1) neurological evidence, showing that pain and pleasure arise in separate brain regions, which can be jointly activated; (2) survey research that looks to how individuals describe their current mental state, finding that words expressing negative affect (words such as "distressed," "upset," "hostile," "irritable," "scared," "afraid," "ashamed," "guilty," "nervous," and "jittery") are surprisingly uncorrelated with words expressing positive affect (words such as "attentive," "interested," "alert," "excited," "enthusiastic," "inspired," "proud," "determined," "strong" and "active"); (3) possible behavioral evidence of the dual activation of positive and negative affective systems; and (4) introspection, namely the possibility of states that are simultaneously painful and pleasurable. Nico Frijda eloquently summarizes the introspective evidence:

---

182. *See, e.g.*, John T. Cacioppo & Gary G. Berntson, *Relationship Between Attitudes and Evaluative Space: A Critical Review, with Emphasis on the Separability of Positive and Negative Substrates*, 115 PSYCHOL. BULL. 401 (1994); Richard J. Davidson, *Anterior Cerebral Asymmetry and the Nature of Emotion*, 20 BRAIN & COGNITION 125 (1992); Ed Diener & Robert A. Emmons, *The Independence of Positive and Negative Affect*, 47 J. PERSONALITY & SOC. PSYCHOL. 1105 (1985); Nico H. Frijda, *Emotions and Hedonic Experience*, *in* WELL-BEING: THE FOUNDATIONS OF HEDONIC PSYCHOLOGY, *supra* note 1, at 190; Peter J. Lang, *The Emotion Probe: Studies of Motivation and Attention*, 50 AM. PSYCHOLOGIST 372 (1995); Richard E. Lucas, Ed Diener & Eunkook Suh, *Discriminant Validity of Well-Being Measures*, 71 J. PERSONALITY & SOC. PSYCHOL. 616 (1996); David Watson, Lee Anna Clark & Auke Tellegen, *Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS Scales*, 54 J. PERSONALITY & SOC. PSYCHOL. 1063 (1988); David Watson & Auke Tellegen, *Toward a Consensual Structure of Mood*, 98 PSYCHOL. BULL. 219 (1985).

> In actual fact, pleasure and pain can coexist at the same moment, and they may do so without annulling each other; masochistic pleasure is a case in point. Simultaneous arousal of both affects leads to complex interactions: the confusing feelings of ambivalence, the experience of conflict, a pungency and excitement added to the pleasure, a sweetness added to the pain, as in nostalgic remembrance. Ambivalence as well as bittersweet experience lead to the hypothesis that pleasure and pain are not the opposite outcomes of one single process, but the outcome of two somewhat independent processes.[183]

How, specifically, might the multidimensionality of affect jeopardize Kahneman's framework for measuring hedonic value? For simplicity, consider the two-dimensional case (the discussion generalizes). Assume that affect is "bivalent": positive and negative affect are not twin poles of a single dimension, but two, orthogonal aspects of experience that can be jointly realized: a moment of experience can possess positive affect, negative affect, or *both*. Kahneman's formal model does not appear to cover this last case. On that model, negative hedonic utilities are assigned to painful mental states, representing that their affective quality is worse than neutrality; and positive hedonic utilities are assigned to pleasurable mental states, representing that their affective quality is better than neutrality. But what sort of numbers should be assigned to bivalent experiences?

Kahneman, indeed, takes very seriously the bivalence objection, and at multiple junctures seeks to rebut it. First, Kahneman argues that bivalent experiences occur seldomly, because the brain's pain and pleasure systems are mutually inhibitory.

> Cacioppo, Gardner, and Berntson (1999) point out that positive and negative affective states are processed by different neural systems and may be activated concurrently. . . . However, the systems are not functionally independent, and there is evidence that they inhibit each other. Lang (1995) has shown, for example, that watching pleasant pictures of food or smiling babies attenuates the startle response to a loud sound, whereas startle is actually enhanced in the presence of disgusting or horrible pictures.[184]

This *empirical* response argues that, even if the "objective happiness" measurement framework is only applicable to a subset of

---

183.   Frijda, *supra* note 182, at 195.

184.   Kahneman, *Experienced Utility*, *supra* note 176, at 682.

experiences—univalent ones—the framework remains a useful tool, because most experiences *are* univalent. However, the assertion of mutual inhibition and, thus, the rarity of bivalence, are empirically contestable.

But Kahneman also offers a second rebuttal to the bivalence objection—one that, if persuasive, would obviate the first. He argues (seemingly) that the brain assigns even bivalent experiences a *net* hedonic value, namely *on balance* positive or negative.

> The bivalent nature of the Good/Bad system is not necessarily incompatible with the notion that most moments can be usefully characterized by a single value on a bipolar Good/Bad dimension. A bivalent system yields a bipolar dimension if the separate mechanisms that mediate Good and Bad are mutually inhibitory or reciprocally innervated *or* if the relevant output of the system is the difference between the levels of activity of the two mechanisms . . . . Davidson (1992) suggested that the brain may compute . . . the difference of the levels of activity in the separate systems that mediate positive and negative affect. He proposed that the [Good/Bad] value corresponds to the difference . . . .[185]

If the claim, here, is that any experience feels *on balance* good or bad, that claim is undermined by the introspection data which Frijda mentions. But perhaps the claim is slightly different, namely that every bivalent experience (although perhaps ambivalent in how it feels) corresponds to some univalent experience for purposes of the observer's ranking of temporally extended experiential profiles. The claim, thus understood, is no longer about the psychology of momentary affect, but rather about how univalent, bivalent, and neutral experiences fit together to determine the well-being value of profiles. But is the claim true? For example, does a sequence of balanced bittersweet moments, wherein pain and pleasure have equal intensity, possess just the same well-being value as a sequence of neutral moments with neither affective system activated?

2. *Do Observers Have the Same Ranking of Hedonic Profiles?* Let us ignore, henceforth, the possibility of bivalent experiences. Let us also place to one side deep questions regarding the accessibility of hedonic (and more generally mental) life, namely whether a subject can ever communicate to an observer what she (the subject) is feeling

---

185. Kahneman, *Objective Happiness*, *supra* note 176, at 8 (emphasis added) (citation omitted) (citing Davidson, *supra* note 182).

like. Assume, instead, that subjects can classify experiences as hedonically good, bad or neutral; can rank experiences with the same valence; can describe experiences using labels or instantaneous utilities; and that these descriptors more or less succeed in communicating to a given observer what the experiences feel like.[186]

Kahneman's formal model assumes a *single* ranking of "profiles," that is, temporally extended hedonic episodes. In effect, he assumes that all observers will converge in ranking any given profile as better, worse, or equally good as any other. But why should this be the case? Kahneman is ambiguous as to whether observers are meant to rank profiles by consulting (a) their preferences (Which profile would I prefer to experience?) or (b) their judgments (Which profile do I think is better for the subject's well-being?). In either event, there is no good reason, conceptual or empirical, to suppose interobserver convergence.

Understand that interobserver convergence is not entailed by the requirement that each observer's $v(.)$ function respect the subject's rating of momentary experiences, nor by the requirement that each observer's ranking of profiles satisfy axioms of time-neutrality, separability, and monotonicity. Two observers might each satisfy all these requirements and yet rank profiles differently.

For example, assume that the subject rates momentary experiences A through C as better than neutral and A better than B better than C. Observer One is indifferent between T hours of each experience and $2T$ hours of the subsequent experience (so between T hours of A and $2T$ hours of B, and between T hours of B and $2T$ hours of C). Observer Two is indifferent between T hours of each experience and $3T$ hours of the subsequent experience. Then Observer One's ranking is represented by the function $v_1(.)$ such that $v_1(A) > v_1(B) > v_1(C) > 0$, $v_1(A)/v_1(B) = 2$, $v_1(B)/v_1(C) = 2$. Observer Two's ranking is represented by the function $v_2(.)$ such that $v_2(A) > v_2(B) > v_2(C) > 0$, $v_2(A)/v_2(B) = 3$, $v_2(B)/v_2(C) = 3$.

Each observer's ranking of profiles (using the sum of duration-weighted $v(.)$ values) satisfies the axioms of time-neutrality, monotonicity, and separability, as well as respecting the subject's feelings. And yet the observers disagree in their ranking of certain profiles. For example, it is easy to see that Observer One will prefer

---

186. Inaccessibility would jeopardize *any* approach that counts mental experience as one component of well-being (including the approach I argue for in the next Section), not merely the "objective happiness" framework.

the profile consisting of two hours of *B* followed by one hour of *C*, over the profile consisting of one hour of *A*, whereas Observer Two will have the opposite preference.

If observers can have different rankings of profiles, Kahneman's framework for arriving at a cardinal measure of momentary hedonic experience collapses. If one observer prefers profile *P* to profile *P**, and another observer has the opposite preference, then clearly there is *no v(.)* function such that the sum of duration-weighted *v(.)* values will *both* assign profile *P* a greater overall value than profile *P**, and assign profile *P** a greater overall value than profile *P*. Consider the case just mentioned in which Observer One prefers two hours of *B* followed by one hour of *C* to one hour of *A*, whereas Observer Two has the opposite preference. It is mathematically impossible to find a *v(.)* such that: $2v(B) + v(C) > v(A)$ *and* $2v(B) + v(C) < v(A)$.

In short, when observers have different rankings of profiles, there will be *no v(.)* function that represents all of their rankings. Rather, the cardinal measurement of hedonic experience will be observer-relative. Observer One will have a cardinal function $v_1(.)$, the sum of whose duration-weighted values will represent *his* ranking of profiles; Observer Two will have a different cardinal function $v_2(.)$, the sum of whose duration-weighted values will represent *his* ranking of profiles; and so on. Policy aggregates such as gross national happiness, requiring a cardinal measure of hedonic quality, will also become observer-relative. For example, gross national happiness might be greater in outcome *x* than *y*, as calculated by adding up the momentary utilities that track Observer One's ranking of profiles; but greater in outcome *y* than *x*, as calculated by adding up the momentary utilities that track Observer Two's ranking of profiles.[187]

3. *Circumventing the Observer in Empirical Implementation.* In his empirical implementation of the "objective happiness" approach, Kahneman has not actually surveyed an observer or observers to rank hedonic profiles and thereby cardinalize hedonic utilities. (Consider, by way of contrast, empirical implementation of the QALY approach to policy analysis, in which observer preferences over health states *are*

---

187. Kahneman at one point suggests that ordinal rather than cardinal momentary utilities will typically be adequate for policy purposes. *See* Kahneman & Riis, *supra* note 176, at 290 ("Except for the rare cases in which cumulative distributions [of moment utility over time] cross, the mean (or the median) of the distribution of moment utility is an ordinal measure of total utility that can be compared across situations, people and populations."). The assertion that the crossing of cumulative distributions will be "rare" is pure speculation.

actually used to cardinalize health values.)[188] Rather, Kahneman has gone directly from subjects' descriptions or ratings of their health states, to conclusions about their "objective happiness."

Circumventing the observers renders the empirical exercise less onerous. It also suppresses difficulties that arise when observers have divergent rankings. But one can ask whether the results of such an exercise have much to do with the account of "objective happiness" that Kahneman defends in his theoretical work.

Consider the study of working women in Texas, which I mentioned in the Introduction.[189] Recall that each respondent was asked to recollect the episodes of her previous day, and to rate each episode on various scales of positive or negative affect, all ranging from values of 0 to 6: for example, how "happy" she felt on a scale of 0 to 6 during the episode, how "warm/friendly" she felt, how "frustrated/annoyed" she felt, or how "depressed/blue" she felt. A measure of the respondent's "positive affect" during the episode was calculated by averaging how she rated the episode on three of the scales (happy, warm/friendly, enjoyment), and her "negative affect" during the episode was similarly calculated by averaging her ratings on six other scales (frustrated/annoyed, depressed/blue, hassled/pushed around, angry/hostile, worried/anxious, criticized/put down).

These positive and negative values were averaged, across subjects, to yield mean affect ratings of different activities. For example, Kahneman found that eating has an average positive-affect rating of 4.34, exercise of 4.31, and watching TV of 4.19. He used such averages not only to compare the affective value of different activities, but also to analyze the diurnal pattern of affect, and to assess whether affect is more influenced by an individual's income or her temperament.

However, the mean affect ratings employed in the Texas study seem quite arbitrary—*arbitrary from the perspective of Kahneman's own account of objective happiness*. Moreover, this is true even if one brackets the problem of interobserver divergence. *If* observers converge—a heroic assumption—then Kahneman's model allows for the assignment of observer-independent cardinal values to momentary experiences: cardinal values, the duration-weighted sums of which correspond to the single ranking of profiles that each and every observer shares. As already explained, these cardinal values

---

188.   *See supra* notes 77–84 and accompanying text.
189.   *See supra* notes 7–8 and accompanying text.

DUKE LAW JOURNAL [Vol. 62:1509

express time-tradeoff judgments. If $v(A) = 3v(B)$, then every observer is indifferent between one hour of *A* and three hours of *B*.

But there is no particular reason to think that the values elicited from subjects via the rating question in the Texas study correspond to these $v(.)$ values. When a subject says that her happiness during commuting was "3" on a 0-to-6 scale, and that her happiness during socializing was "5" on a 0-to-6 scale, is she using these numbers to express her time-tradeoff judgments? Is she saying that she is indifferent between *T* hours commuting, and $3/5T$ hours socializing? Why assume that *this* is what she means? Perhaps she means the numbers only to have ordinal significance: "5" for socializing and "3" for commuting just means that socializing is more pleasurable than commuting. Or perhaps she intends the numbers to do more than represent the hedonic ordering of experiences, but to communicate some feature of those experiences, or of her ranking of temporal or probabilistic bundles of experiences, *other* than her time-tradeoff judgments.[190]

In his most recent empirical work, Kahneman employs a different approach (the "U-index") to measure the hedonic value of moments.[191] Subjects are still asked to assign numerical ratings to a given episode on different scales of positive and negative affect. But the episode is then assigned a value of 1 (if the highest rating was for a scale of negative affect), otherwise 0. Kahneman claims that this approach is purely ordinal.[192] That claim is somewhat suspect.[193] A

190. The numbers might be cardinal, but in the sense of (1) representing the respondent's ranking of lotteries over experiences, as per expected utility theory; (2) representing the respondent's judgments regarding the hedonic differences between experiences; (3) deriving from a multiattribute function that aggregates subutility functions corresponding to the multiple dimensions of hedonic experience; or (4) communicating a primitive, cardinal measure of the felt intensity of the experiences. On this point, note too that QALY values elicited via explicit time-tradeoff questions need not—and in practice, do not—correspond to QALY values elicited through a direct rating question or the standard-gamble question. *See supra* note 77.

191. Krueger et al., *supra* note 4, at 9.

192. *See id.* at 19 (stating that "[t]he U-index is an ordinal measure *at the level of feelings*" because the classification of an episode as unpleasant or pleasant "relies purely on an *ordinal ranking* of the feelings within each episode").

193. In the theoretical presentation of the "objective happiness" framework, it is assumed that momentary experiences can be compared to the neutral level, and that experiences within each hedonic domain (positive or negative) can be compared to each other. So pains are ordered, pleasures are ordered, but pains and pleasures are not necessarily ordered vis-à-vis each other *except* in the sense of being categorized as worse or better than neutral. The U-index procedure assumes that the intensity of any positive or negative experience can be compared to the intensity of any other positive or negative experience. *See id.* at 19 n.13. It is much more contestable whether subjects can make such comparisons.

truly ordinal, and more direct, approach would be to ask subjects a binary question, namely to characterize each episode as, on balance, worse or better than neutral experience—and to assign it the number 1 in the first case, 0 in the second. But perhaps this direct, binary, survey would be unreliable, and the actual U-index methodology can be seen as a reliable proxy for it.

In either event, assigning experiences a number of 1 (for unpleasant) or 0, and then adding up or averaging these binary numbers, is a very crude way to assess overall or average hedonic value of a person or a group. We lose all information about the relative intensity of hedonic experiences. Jim's profile of experiences on Tuesday might be assigned a larger average U-index value than Jim's profile today, even though all observers would converge in preferring the Tuesday profile or in judging it better.[194] The U-index study, like the Texas study, avoids the empirical encumbrance of actually introducing observers, but it does so at the cost of results that seem arbitrary by the lights of Kahneman's own theory of how to measure affects.

4. *Temporal Separability*. Kahneman's measurement scheme assumes the temporal separability of hedonic value. Roughly, this means that the sequencing of momentary utility does not affect the ranking of temporally extended episodes. More precisely, it says that if two episodes have the same hedonic value during some moment or moments, the ranking of the episodes does not depend on what particular value that is.

Kahneman's argument for temporal separability runs as follows:

> The ordering of experiences can affect the utility they confer. For example, a strenuous tennis game and a large lunch yield a better experience in one order than in the other because the enjoyment of the tennis game is sharply reduced when it follows lunch. The condition of separability, which states that the contribution of an element to the global utility of the sequence is independent of the elements that preceded and followed it, is often violated when the sequences are described in terms of physical events such as lunch and a tennis game. In a moment-based treatment of total utility, however, the elements of the sequence that is to be evaluated are not events but rather moment utilities associated with events.

---

194. Imagine that Jim on Tuesday has more negative moments than Jim today, but that his positive moments on Tuesday—when they occur—are intensely positive, whereas nothing so joyful happens to him today.

> Because *all* the effects of the order of events are already incorporated into moment utilities, the order of these moment utilities no longer matters.[195]

However, Kahneman's insistence that "the order of . . . moment utilities no longer matters" is sheer ipse dixit. We are back to the issue of observer convergence and divergence. Might not *some* observers judge/prefer lives that trend upwards in terms of momentary hedonic value, just as some judge/prefer lives that trend upwards in terms of income or health?[196]

5. *Integrating Hedonic Value with the Nonhedonic Components of a Good Life*. Hedonic value is *one* determinant of well-being, but not the only determinant. It is implausible to claim that an individual's welfare is *solely* constituted by the affective quality of her experiences. First, such a claim would amount to a kind of experientialism about well-being; and, as I have already discussed, experientialism is quite controversial.

Second—and independent of the debate about experientialism—we should reject hedonism about mental well-being, that is, the position that hedonic value is the only welfare-relevant aspect of an individual's mental life.[197] Various examples can be constructed to drive home this point, but consider just one. Cheery is upbeat, but forgets much of what happens to him, and his stock of propositional knowledge is pretty mediocre. Grumpy goes through daily life in an affective state that is at or slightly below neutral, but he is keenly focused on the sights and sounds around him, can recollect his past in rich detail, and has educated himself in various fields. Assume, further, that Cheery and Grumpy are more or less the same in their objective characteristics (income, health, job status, social life). Is it clear that Cheery lives a better life than Grumpy? Hardly.

Indeed, hedonism about mental well-being is regularly rejected by SWB researchers themselves. Ed Diener, a leading figure, has consistently defined subjective well-being as a hybrid of affect and life

---

195. Kahneman, *Experienced Utility*, *supra* note 176, at 678.

196. *See* ADLER, *supra* note 27, at 419–20; *see also* Loewenstein, *supra* note 42, at 100–02 (challenging an approach that measures experiential quality by summing moment utilities, and noting both that peak negative and positive experiences have an importance that is disproportionate to their duration and that individuals encode experiences in terms of "episodes" rather than moments).

197. *See, e.g.*, Loewenstein & Ubel, *supra* note 19 at 1801–04; Loewenstein, *supra* note 42 at 94–96.

satisfaction.[198] Paul Dolan, Richard Layard, and Robert Metcalfe, in a report laying the groundwork for large-scale gathering of SWB statistics by the government of the United Kingdom, recommend that an individual's SWB be assessed using three different kinds of measures: measures of affect, measures of life satisfaction, and "eudaimonic" measures, the last assessing how "rewarding" or "worthwhile" individuals perceive their activities to be.[199] Kahneman himself has reversed course and now concedes:

> [E]xperienced and evaluated well-being should both be measured, and . . . the measures should be explicitly separated. Contrary to a position that one of us espoused earlier (Kahneman 1999), measures of evaluated well-being are not simply flawed indicators of objective happiness (experienced well-being). Evaluation and memory are important on their own . . . because people care deeply about the narrative of their life.[200]

However, Kahneman continues to argue that his "objective happiness" framework is a good tool for incorporating hedonic value into policy assessment. The hedonic utility of the positive and negative affective states experienced by each individual should be measured separately from whatever nonhedonic attributes are welfare-relevant;[201] and these hedonic utilities should be aggregated or averaged across persons and moments, as in the Texas and U-index studies.

But there are two difficulties—one obvious, the second more subtle—with this position. First, the "objective happiness" framework, itself, offers no guidance in balancing the hedonic and nonhedonic aspects of well-being. Imagine that we predict that a policy will increase the average "objective happiness" of some population, but reduce some measure of their average nonhedonic well-being. Is the policy worthwhile, all things considered?

The more subtle point is that the very assignment of hedonic utilities to individual moments—at least as Kahneman conceptualizes it—assumes the separability of hedonic and nonhedonic attributes. Remember his formal model. The observer considers temporally extended episodes ("profiles"), characterized in terms of the affects

---

198. *See, e.g.*, Diener et al., *supra* note 1, at 277.

199. DOLAN ET AL., *supra* note 17, at 6–9.

200. Kahneman & Riis, *supra* note 176, at 289 (citing Kahneman, *Objective Happiness*, *supra* note 176).

201. *See id.*

experienced by the subject during each moment, and ranks these—yielding a cardinal measure of momentary hedonic experience, what I referred to above as the $v(.)$ value of that experience. But—now that Kahneman has, properly, conceded the welfare-relevance of nonhedonic attributes—the insistence that any given observer *has* such a ranking becomes problematic.

To be sure, the observer will have a *conditional* ranking of hedonic profiles. Assume that N is a possible profile of *non*hedonic, welfare-relevant, attributes that a subject might possess. N includes both nonhedonic mental attributes (an individual's memory, perceptions, sense of satisfaction, and so forth) and whatever nonmental attributes are seen to be welfare-relevant (for example, health). H is one possible hedonic profile, that is, a sequence of hedonic attributes, $H*$ a second possible hedonic profile, $H**$ a third, and so forth. Then the observer *will* be able to rank NH versus $NH*$ versus $NH**$. We can call this the ranking of hedonic episodes *conditional* on nonhedonic profile N. Similarly, if N^+ is a different specification of nonhedonic attributes, the observer *will* be able to rank N^+H versus N^+H* versus N^+H**. We can call *this* second ranking the ranking of hedonic episodes conditional on nonhedonic profile N^+.

But it is a further question whether the observer's ranking of hedonic episodes is *invariant* to nonhedonic attributes, namely, that if the observer has a particular ranking of hedonic episodes conditional on N, then she has the very same ranking conditional on N^+, on N^{++}, and every other nonhedonic profile. Why believe that invariance holds true? Might it not be possible that the observer prefers NH to $NH*$ to $NH**$, but N^+H** to N^+H to N^+H*? Invariance precludes *interaction effects* between affective content, on the one hand, and anything else occurring in someone's life on the other. The preclusion of such interaction is highly problematic. For example, who would want to experience a mental life in which the pairing of affects and memories is random, as opposed to being appropriate to what is remembered? Similarly, insofar as health is an intrinsic determinant of well-being, why think that the ranking of temporally extended sequences of pains and pleasures is invariant to the sequence of health states?

It is tempting to respond that the possible interaction between hedonic and nonhedonic attributes is a nuance which policy tools, necessarily crude and approximate, may need to ignore. But this response misunderstands the objection. *Once* we have assigned

hedonic utilities to moments, we may (for practical reasons) need to ignore their correlation with nonhedonic attributes, in evaluating policies. However, we still need some rationale for this assignment. We still need some theory of *what* is being measured. Kahneman's theory is that hedonic utilities represent the preferences/judgments of an observer, ranking temporally extended hedonic episodes. But— once interaction effects are allowed into the picture—the observer may *have* no such ranking. If you ask her, "Do you prefer *H* to *H** or *H** to *H*?" she may respond: "The question is meaningless. I have various conditional preferences/judgments regarding the episodes, but no unconditional preference/judgment. To ask me for such a preference/judgment is like asking whether I prefer a long-sleeved or a short-sleeved shirt, to which I'll answer, "The first in cold weather, the second in warm.""

C.   A Better Approach?

In this Section, I very briefly sketch a different methodology than the "objective happiness" framework for incorporating information about experiential quality into policy analysis. I lack space to elaborate this methodology at length, but at least I can suggest how the limitations in Kahneman's framework might be remedied.

A preference-based view of well-being allows for an individual's happiness, or some other aspect of her experiential quality, to be one of the fundamental arguments for her preferences. Indeed, it would be absurd for the preferentialist to insist that only nonmental properties can be intrinsic determinants of well-being! However, the simplified utility functions adopted in classical economic theory and in much empirical work have ignored experiential arguments. Instead, preference utility is often modeled as a function of one or more nonmental attributes, quintessentially income, bundles of commodities, health, leisure, or the physical state of the environment.

This sort of simplified model of preferences, while more tractable, has a downside when coupled with policy tools such as cost-benefit analysis. Policy impacts on happiness, and thus on well-being, cannot be directly measured. Instead, it must be assumed (plausibly) that policies causally affect happiness via changes to individuals' nonmental attributes and (less plausibly) that individuals can accurately predict how changes in nonmental attributes affect their happiness and thus that their intrinsic preferences for happiness are

ADLER IN PRINTER PROOF (DO NOT DELETE)   4/15/2013 2:56 PM

fully captured by the instrumental component of their preferences for nonmental attributes.[202]

A richer model of preferences explicitly includes experiential properties. A given attribute bundle *A* is characterized as $A = (M, N)$, where *M* describes one or more types of mental attributes, and *N* nonmental attributes. For short, call this a "hybrid bundle." Individual *i*'s preferences over hybrid attribute bundles are inferred from stated-preference surveys or, perhaps, behavioral evidence. Individual *i*'s *ordinal* preference-utility function for hybrid bundles captures his ranking of these bundles. That ordinal preference-utility function takes the form $u_i(M, N)$. If *income* is included among the nonmental attributes, this ordinal preference-utility function is sufficient to determine individual *i*'s willingness to pay/accept for changes in the mental attributes.

Information about individual WTP/WTA for experiential changes allows for the direct incorporation of such changes in policy modeling. For example, imagine that a costly proposed regulation, lowering the ambient level of some feared environmental toxin, will have the benefit of reducing deaths and injuries *and* reducing anxiety about the toxin. Information about individual WTP/WTA for fear and anxiety allow us to undertake a cost-benefit test of the regulation that explicitly identifies fear-reduction as a separate "good" additional to physical harm-avoidance.

An individual's *cardinal* utility function for hybrid bundles would be determined by her ranking of bundles plus further facts about her preferences (for example, her ranking of bundle lotteries or her time-tradeoff preferences). Such cardinal utility functions are *not* required by cost-benefit analysis but might be required by other policy tools (for example, a GDP-like measure that aggregates/averages preference-utility across persons, with preference-utility in turn dependent on both experiential and nonexperiential attributes).

The approach now under discussion might seem utopian (or dystopian, for skeptics about policy analysis) but in fact has some precedent in existing empirical work. A small literature *does* try to estimate individual WTP/WTA for fear and anxiety (with application both to environmental regulation and to policing strategies that

---

202. The literature on errors in individual affective forecasting makes this latter premise less plausible. *See, e.g.*, Kahneman & Sugden, *supra* note 121, at 168–73.

mitigate the fear of crime).[203] The QALY literature looks at individual preferences over health states. The QALY number assigned to a health state (between 0 and 1) captures such preferences, for example, time-tradeoff preferences as between living a period of time in perfect health and living longer in a poorer health state. The health states thus valued sometimes encompass mental, not just physical health. It is not unusual to find QALY values for conditions such as depression, pain, or chronic anxiety.[204]

Estimating individuals' preference-utility for hybrid bundles via standard preference-elicitation techniques (stated-preference surveys or inference from behavioral evidence) is to be sharply distinguished from the methodology criticized in Part II: using SWB surveys as evidence of preference utility, that is, taking an individual's stated life satisfaction when in possession of bundle *A* as his preference utility for *A*. What is being recommended is *not* that an SWB number be employed as a proxy *for* preference utility, but rather that an individual's experiences be included among the determinants *of* preference utility.

Indeed, this approach does not essentially rely upon the assignment of numbers to mental states. In a stated-preference survey, the experiential component of a hybrid bundle could be described via a number ("imagine experiencing pain that you might rate at 4 on a scale of 1 to 5"), but it could also be described qualitatively—and indeed a qualitative description might be more successful in accurately communicating to respondents what the state feels like.

The approach also has key advantages over Kahneman's "objective happiness" framework. First, although that framework seeks only to quantify the well-being impact of affective states, in a hedonic utility number, the "hybrid bundle" approach is more generic. If $A = (M, N)$, with M mental attributes, M itself might be unidimensional or multidimensional and, in either event, might include affective states, memory, cognition, sense of satisfaction or purpose, and so forth. Second, the "objective happiness" framework

203. *E.g.*, Matthew D. Adler, *Fear Assessment: Cost-Benefit Analysis and the Pricing of Fear and Anxiety*, 79 CHI.-KENT L. REV. 977, 1024–30 (2004); Paul Dolan, Graham Loomes, Tessa Peasgood & Aki Tsuchiya, *Estimating the Intangible Victim Costs of Violent Crime*, 45 BRIT. J. CRIMINOLOGY 958, 959–64 (2005).

204. *See* Adler, *supra* note 203, at 979–80, 1043–52; Paul Dolan & Tessa Peasgood, *Estimating the Economic and Social Costs of the Fear of Crime*, 47 BRIT. J. CRIMINOLOGY 121, 126 (2007).

DUKE LAW JOURNAL   [Vol. 62:1509

may be undermined by the bivalence of pain and pleasure—the fact that positive and negative affect may really be two, separate, dimensions of experience—but the hybrid-bundle approach is perfectly compatible with affective bivalence. Third, the hybrid-bundle approach is not committed to any kind of *separability* with respect to an individual's ranking of bundles: temporal separability, the separability of hedonic value from other aspects of experiential quality, or the separability of hedonic and nonmental attributes. Fourth, the approach provides a basis for making all-things-considered policy judgments, integrating information about policy impacts on experiential quality with other sorts of impacts. If we know someone's preference utility for different hybrid bundles, then we know how she makes tradeoffs between the experiential and nonexperiential components of these bundles. The *governmental* choice between multiple policies, characterized by different kinds of experiential and nonexperiential effects, can then be a function of the totality of individual tradeoffs—as operationalized via cost-benefit analysis (the sum of individual WTP/WTA amounts), or perhaps in some other way.

Fifth, at least insofar as this approach seeks only to infer individuals' ordinal preference-utility functions—which is all that is required for purposes of CBA and the determination of WTP/WTA amounts—the approach does *not* presuppose that individuals have the same preferences.[205] The ranking of hybrid bundles can vary from individual to individual. By contrast, as we have seen, Kahneman's "objective happiness" framework makes the implausible presupposition that "observers" will have the very same ranking of temporally extended hedonic episodes ("profiles").[206]

The elicitation of preference utility for hybrid bundles presupposes that individuals have well-behaved preferences over such bundles—if not initially, then at least after debiasing and information-

---

205. Let *x* and *y* be two outcomes. In order to determine what a given individual is WTP/WTA for the move from *x* to *y*, all we need to know, in principle, is how he orders attribute bundles (including both income and whatever non-income attributes are specified in the outcomes). This ordering is captured by his ordinal utility function. Moreover, individuals *i* and *j* might have different rankings of attribute bundles, and suitable estimation techniques (such as stated-preference surveys) are robust to such heterogeneity. *See supra* Part II.C.

   Admittedly, if we were to compare *x* and *y* by using a *cardinal* utility function for attribute bundles, heterogeneity in utility functions *would* become a problem. Comparing the sums or averages of bundle utilities using one method for assigning cardinal utilities could yield a different result than comparing those sums or averages using a different method.

206. *See supra* notes 186–187 and accompanying text.

provision. Preference skeptics will deny this, pointing to violations of rationality conditions[207]—and this challenge *is* a serious challenge to the "hybrid bundle" approach. But, as I have just reminded the reader, Kahneman's theory for assigning hedonic utilities to moments *also* depends upon an observer having a preference/judgment ranking of hedonic profiles. This observer is suppressed in Kahneman's empirical work—with consequent difficulties that I elaborated upon in discussing the Texas and U-index studies. It is ironic that Kahneman, in some scholarship, comes across as an ardent preference skeptic.[208] If actual individuals (even after debiasing) are incapable of satisfying rationality conditions requisite for holding a preference, then who is the observer whose well-behaved ordering of profiles is supposed to satisfy such conditions? In short, Kahneman's "objective happiness" framework is no more immune from preference skepticism than the hybrid-bundle approach.

CONCLUSION

Enthusiasm about the policy role of SWB surveys is premature. Why think that the number which someone assigns to her momentary or overall happiness, life satisfaction, positive or negative affect, or some other aspect of her experiential state offers real help in evaluating governmental policies? Two different answers to this question need to be teased apart. One says that a higher self-rated degree of life satisfaction shows that the respondent's preferences are more fully realized. In short, SWB surveys evidence *preference utility*. But the evidence would seem to be pretty poor. Preference and scale heterogeneity hamper the use of self-rated life satisfaction to make inferences about preference utility. Even if all respondents share the same underlying preferences and utility function, someone's answer to an SWB survey may well be skewed by evaluation error or miscommunication. This number may well be an inaccurate and, indeed, statistically biased indicator of the degree to which her life-circumstances realize her preferences.

---

207. Either the minimal conditions required to have a preference at all, or the conditions requisite for the preference to have normative "bite." *See supra* notes 35–48.

208. *See* Dolan & Kahneman, *supra* note 56, at 215–16 (questioning the usefulness of measures of "decision utility," that is, preference utility, as a basis for valuing health states, because individuals' preferences regarding health are inevitably biased to some extent); Kahneman et al., *supra* note 121, at 228–29 (claiming that "people are better described as having attitudes than preferences," with attitudes "lack[ing] some of the essential properties that economic theory requires of preferences").

Stated-preference surveys dominate SWB surveys *as evidence of preference-realization*. Anomalies using the stated-preference format suggest the importance of debiasing preferences—rendering them rational and well-informed. Perhaps debiasing is fruitless. But that would show that government policy choice must (somehow) find a normative foundation other than individuals' preferences, and not that preferences should be inferred via the SWB technique. Skepticism about the rationality of preferences hardly advances the PR (preference-realization) defense of SWB surveys.

The second answer to the "why" question takes a different tack, suggesting that a happiness, affect, or life-satisfaction rating is a measure of *experiential quality*. Thus goes the EQ defense of SWB surveys. Kahneman's "objective happiness" framework—using SWB surveys focused on momentary hedonic quality—is an important first step in developing a policy-relevant measure of experiential quality. Kahneman does not argue that well-being and good experiences are equivalent—but rather, much more plausibly, that good experiences are one important *aspect* of well-being.

However, a close examination of the "objective happiness" framework suggests significant limitations. The framework purports to cardinalize momentary hedonic utilities by appealing to an "observer's" ranking of temporally extended hedonic episodes, but presupposes—without justification—that observers have the same ranking, and that these rankings are separable from nonhedonic attributes. In empirical implementation, Kahneman has suppressed the observer and, most recently, abandoned any attempt at cardinalization—via a "U-index" that merely reports the fraction of time that individuals spend in an affectively unpleasant state. This is a crude measure of hedonic quality (let alone the nonhedonic aspects of experiential life, such as memory or a sense of meaning), because it does not tell us about the intensity of individuals' affective states.

It remains unclear whether SWB surveys—asking for a numerical rating of experiences—should be the central tool for incorporating information about experiential quality into policy analysis. At least in principle, a different approach, more closely continuous with traditional cost-benefit analysis, is available: namely, to use revealed or stated-preference evidence to infer individuals' preferences over "hybrid bundles," comprising both experiential and nonexperiential attributes. SWB surveys are at most an ancillary component of this approach. Its central focus is inferring preference

utility, with experiential attributes merely one entry in the utility function.

Much more work remains to elaborate both this approach and frameworks (such as "objective happiness") that revolve around SWB surveys. In undertaking this effort, scholars should exercise caution, taking care not to muddy their concepts—taking care to understand that well-being need not reduce to good experiences, that individuals can have intrinsic preferences for aspects of their lives other than their mental states, and that someone's perceived degree of happiness or life satisfaction can diverge from her true preference utility.