

# Toward National Well-Being Accounts

By DANIEL KAHNEMAN, ALAN B. KRUEGER, DAVID SCHKADE, NORBERT SCHWARZ,  
AND ARTHUR STONE\*

Economists have traditionally eschewed direct measures of well-being on methodological grounds: the private nature of experience and the discomfort of making interpersonal comparisons. Instead, income is often used as a proxy for opportunities and well-being. If people are not fully rational, however, their choices will not necessarily maximize their *experienced utility*, and increasing their opportunities will not necessarily make them better off (Kahneman, 1994; Cass R. Sunstein and Richard Thaler, 2004). Direct measures of experienced utility become particularly relevant in a context of bounded rationality.

Furthermore, advances in psychology and neuroscience suggest that experienced utility and well-being can be measured with some accuracy (Kahneman et al., 1999). Robust and interpersonally consistent relationships have been observed between subjective measures of experience and both specific measures of brain function and health outcomes. In part because of these findings, economic research using subjective indicators of happiness and life satisfaction has proliferated in recent years (see Bruno Frey and Alois Stutzer [2002] for a survey). Most work on well-being uses a question on overall life satisfaction or happiness. We suggest an alternative route based on time budgets and affective ratings of experiences.

## I. Plausible and Puzzling Findings of Well-Being Research

Numerous studies have established that life satisfaction is weakly correlated with income and with religiosity, but uncorrelated with either

education or climate. Minnesota, for example, is among the happiest states. The function that relates satisfaction to age is U-shaped: reported happiness rises with age from age 45 to 70, controlling for health. Life satisfaction is low among the unemployed and is affected by life events such as marriage, divorce, and bereavement. People who describe themselves as happy or as satisfied with their health are likely to be extraverted, sociable, and optimistic. They show a characteristic pattern of electrocortical activity, with greater activity in the left than in the right prefrontal cortex (Richard J. Davidson, 2003; H. L. Urry et al., 2004). They have a stronger response to an influenza vaccine and recover more quickly from controlled wounds (J. K. Kiecolt-Glaser et al., 2002; S. Cohen et al., 2003).

The list of plausible results is long, but research using the standard measures of well-being has also produced two major puzzles (Ronald Inglehart and Jacques-René Rabier, 1986): (i) surprisingly small effects of circumstances on well-being (e.g., income, marital status, etc.); (ii) surprisingly large differences in the level of life satisfaction in various countries.

The most remarkable finding in the well-being literature is the extent to which people *adapt* to circumstances, even extreme circumstances. P. Brickman et al. (1978) reported that after a period of adjustment lottery winners were not much happier than a control group, and paraplegics were not much unhappier. Data from the German Socio-Economic Panel indicate that the effects on life satisfaction of both marriage and widowhood largely dissipate within three years of the event (Richard E. Lucas et al., 2003). R. A. Easterlin (1995) finds that average self-reported happiness did not increase in Japan from 1958 to 1987, although real income increased fivefold.

Findings of adaptation are robust, but open to multiple interpretations. Brickman and D. T. Campbell (1971) proposed a *hedonic treadmill* hypothesis: people adapt to situations that are initially pleasant or unpleasant, much as they

\* Kahneman and Krueger: Woodrow Wilson School of Public and International Affairs, Princeton University, Princeton, NJ 08544; Schkade: Department of Management, McCombs School of Business, University of Texas, Austin, TX 78712; Schwarz: Institute for Social Research, University of Michigan, Ann Arbor, MI 48106; Stone: Department of Psychiatry and Behavioral Science, School of Medicine, Stony Brook University, Stony Brook, NY 11794. Krueger is also affiliated with the National Bureau of Economic Research.

adapt to a warm bath. The pleasure or pain evoked by a new situation declines in intensity and is eventually replaced by neutral feelings. Kahneman et al. (1999) observed that mean-reversion is also compatible with the hypothesis of an *aspiration treadmill*: pleasure or pain might persist, but the evaluation of these experiences is relative to expectations, and expectations eventually adjust. On this hypothesis, global reports of subjective well-being exaggerate the amount of hedonic adaptation that actually occurs. The ambiguity can only be resolved by measuring the hedonic quality of experience separately from expectations.

The second puzzle is the consistent finding of large differences in reports of life satisfaction across seemingly similar countries. For example, 64 percent of the Danes described themselves as “very satisfied” with their lives in a Eurobarometer survey, but only 16 percent of the French did so. The difference between the French and the Danes is more than twice as large as the difference between the employed and unemployed in either country. Across 63 countries included in the World Values Survey, the standard deviation of country means of overall satisfaction is 1.12, more than half of the average standard deviation of individuals within countries (2.21). These differences appear implausibly large, and they raise additional doubts about the validity of global reports of subjective well-being, which may be susceptible to cultural differences in the norms that govern self-descriptions (Alex Inkeles, 1993; Ed Diener, 2000; Diener and Eunkook M. Suh, 2000).

## II. Subjective versus Objective Aggregation and Other Potential Biases

F. Y. Edgeworth (1881) imagined a “hedonimeter,” which continuously records an individual’s utility (in Jeremy Bentham’s sense of the term as momentary positive or negative feelings). Happiness is defined by the integral of utility over time. Kahneman et al. (1997) provide a formal analysis of the conditions under which global judgments of the total utility of extended outcomes will satisfy temporal integration. That paper also reviews experimental research demonstrating that individuals’ global retrospective assessments of their experiences consistently violate the logic of temporal integration. In particular, global subjective judgments

of episodes generally overweight experiences that are either extreme or recent, and assign little or no weight to the duration of an experience. People are apparently unable to produce an accurate and unbiased evaluation of experiences that extend over time.

The life satisfaction and happiness questions that are used in well-being research request the type of global assessment that people perform poorly on in the psychological laboratory. Experimental variations of surveys have shown that many irrelevant factors affect these evaluations. Thus, reports of life satisfaction are influenced by manipulations of current mood and of the immediate context, including earlier questions on a survey that cause particular domains of life to be temporarily salient (N. Schwarz and F. Strack, 1999). Satisfaction with life and with particular domains (e.g., income, work) is also affected by comparisons with other people and with past experiences (Andrew E. Clark, 2003). The same experience of pleasure or displeasure can be reported differently, depending on the standard to which it is compared and the context.

In summary, global subjective evaluations of one’s life are unlikely to provide an accurate representation of the concept of utility that Edgeworth proposed. Discrepancies will arise because the durations of experiences are not adequately weighted in global assessments, and because these assessments are unduly influenced by the immediate context and by irrelevant standards of comparison. To overcome these biases we need measures of well-being that have the following characteristics: (i) they should represent actual hedonic and emotional experiences as directly as possible; (ii) they should assign appropriate weight to the duration of different segments of life (e.g., work, leisure, etc.); (iii) they should be minimally influenced by context and by standards of comparison.

## III. Experience Sampling and the Daily Reconstruction Method

The Experience Sampling Method (ESM) collects information on individuals’ experiences in real time in their natural environments (Mihalyi Csikszentmihalyi, 1990; Stone and S. Shiffman, 1994). ESM is intended to overcome problems inherent in global satisfaction questions, namely, imperfect recall and duration

neglect. It is therefore the current gold standard for measurement of well-being in the Edgeworth tradition. ESM is carried out by supplying subjects with an electronic diary (e.g., a specially programmed palm pilot) that beeps at random times during a day and asks respondents to describe what they were doing just before the prompt. The electronic diary also asks respondents to indicate the intensity of various feelings (e.g., happy, frustrated/annoyed, etc.). These data may be averaged to produce a metric reflecting actual daily experience. ESM appears to meet the principal requirements for a measure of well-being that reflects an integration of immediate experience. However, ESM is not a practical method for national well-being accounts (NWBA): (i) it is impractical to implement in large samples; (ii) the rate of nonresponse may be unacceptable for some activities; (iii) infrequent activities are only rarely sampled.

Fortunately, data collected from ESM can be reasonably approximated by other more practical methods. One alternative we developed is the Daily Reconstruction Method (DRM). The DRM asks respondents to fill out a diary corresponding to events of their previous day. Next, respondents describe each episode by indicating: (i) when the episode began and ended; (ii) what they were doing; (iii) where they were; and (iv) whom they were with. To ascertain how they felt during each episode on selected affect dimensions, respondents were asked to fill out the box in Table 1 for each episode. Note that responses are anchored at "not at all," a natural zero point that is likely to have a common and stable meaning for respondents.

The DRM involves a retrospective report on an emotional state, but the procedure was designed to achieve accurate recall, by directing respondents to retrieve specific episodes from memory. The method appears to have been successful: it reproduced a complex pattern of diurnal variation in tiredness and in positive and negative affect, which had previously been obtained in an ESM study (see Kahneman et al., 2003). Data collected from ESM or DRM can be used to characterize the average affective experience that people perceive during particular situations. We use the term "situation" to refer to features of an episode: when, what, where, and who with. Table 2 summarizes the mean affect ratings for selected activities. Net

TABLE 1—THE BOX FILLED OUT BY RESPONDENTS FOR EACH EPISODE IN THE DRM

How did you *feel* during this episode?

*Please rate each feeling on the scale given. A rating of 0 means that you did not experience that feeling at all. A rating of 6 means that this feeling was a very important part of the experience. Please circle the number between 0 and 6 that best describes how you felt.*

	Not at all				Very much		
Happy.....	0	1	2	3	4	5	6
Frustrated/annoyed.....	0	1	2	3	4	5	6
Depressed/blue.....	0	1	2	3	4	5	6
Hassled/pushed around....	0	1	2	3	4	5	6
Warm/friendly.....	0	1	2	3	4	5	6
Angry/hostile.....	0	1	2	3	4	5	6
Worried/anxious.....	0	1	2	3	4	5	6
Enjoying myself.....	0	1	2	3	4	5	6
Tired.....	0	1	2	3	4	5	6

affect is defined as the average of the positive adjectives less the average of the negative adjectives, for individuals engaged in each activity. If an episode involved more than one activity, it enters more than one time, so total hours in a day are not constrained to sum to 24; for NWBA, it would be desirable to either apportion multiple activities that occur in an episode or restrict attention to the focal activity. The sample consists of 909 working women in Texas, and the data are described in more detail in Kahneman et al. (2003). Notice that commuting to and from work and working score relatively low, while leisure activities score high, as expected.

Easier ways of collecting the same type of information also appear to be possible. In ongoing work we have experimented with asking questions about feelings associated with particular events, such as the last episode of commuting to work. We call this the Event Recall Method, or ERM for short. We collected ERM data for another 504 working women in Texas. For most activities, the ERM and DRM yielded insignificant differences. ERM has the important advantage of being easy to administer in a telephone survey. Notice, however, that the selection of who participates in activities, and for how long, differs in ERM and DRM, which would affect the results if heterogeneous preferences lead to very different time allocations across people.

Interestingly, a study of kidney dialysis patients and matched controls using ESM (Jason Riis et

TABLE 2—MEAN NET AFFECT BY ACTIVITY

Activity	Percentage of sample	Time spent (hours)	Net affect
Intimate relations	11	0.21	4.74
Socializing after work	49	1.15	4.12
Dinner	65	0.78	3.96
Relaxing	77	2.16	3.91
Lunch	57	0.52	3.91
Exercising	16	0.22	3.82
Praying	23	0.45	3.76
Socializing at work	41	1.12	3.75
Watching TV	75	2.18	3.62
Phone at home	43	0.93	3.49
Napping	43	0.89	3.27
Cooking	62	1.14	3.24
Shopping	30	0.41	3.21
Computer at home	23	0.46	3.14
Housework	49	1.11	2.96
Childcare	36	1.09	2.95
Evening commute	62	0.62	2.78
Working	100	6.88	2.65
Morning commute	61	0.43	2.03

Notes: Net affect is the average of three positive adjectives (enjoyment, warm, happy) less the average of five negative adjectives (frustrated, depressed, angry, hassled, criticized). All the adjectives are reported on a 0–6 scale, ranging from “not at all” to “very much.” The “time spent” column is not conditional on engaging in the activity. The sample consists of 909 employed women in Texas.

al., 2003) and a study of teachers in exemplary and failing schools using DRM (Kahneman et al., 2003) both find evidence of adaptation, lending support for the hedonic treadmill interpretation.

#### IV. Time-Based National Well-Being Accounts

We return to Bentham and Edgeworth’s notion that utility is the integral of the stream of pleasures and pains associated with events over time. A simple formulation is that utility is time separable. Write an individual’s utility in discrete time as  $U_i = \sum_j h_{ij} \mu_{ij}$ , where  $h_{ij}$  is the amount of time individual  $i$  is engaged in situation  $j$  (e.g., washing the dishes with one’s spouse) and  $\mu_{ij}$  is the net affective experience during situation  $j$ .

A measure of national well-being (WB) therefore is

$$(1) \quad \text{WB} = \sum_i \sum_j h_{ij} \mu_{ij} / N$$

where  $N$  is the population size. Notice that (1) can be written as:  $\sum_j H_j \bar{\mu}_j + \sum_i \sum_j h_{ij} (\mu_{ij} - \bar{\mu}_j) / N$ ,

where  $H_j$  is the average of  $h_{ij}$  over people and  $\bar{\mu}_j$  is the average net affect experienced during situation  $j$ . In our data, time spent on an activity is virtually uncorrelated with net affect across people ( $r = 0.01$ ), so NWBA can be measured by

$$(2) \quad \text{WB}' = \sum_j H_j \bar{\mu}_j.$$

This equation has the advantage that time use and affect can be from separate surveys.<sup>1</sup>

To compute equation (1), net affect and time use can be collected from DRM. For (2),  $\bar{\mu}_j$  can be collected from ERM (or DRM), and  $H_j$  from a separate survey, such as the Bureau of Labor Statistics’ new monthly American Time Use Survey.

There are, of course, many assumptions underlying this formulation. We must assume: that affective experiences can be compared across people; that net affect provides a cardinal measure of utility; utility is time separable; and that a simple measure of net affect represents the utility of an experience. In addition to these conceptual hurdles, there are several practical problems as well: the situations that are relevant for well-being must be identified (what goes into  $j$ ); the allocation of time must be measured; data on net affect for a representative sample in different situations must be collected; and the adjectives that go into defining affect must be specified. The question is not whether (1) provides a perfect measure of well-being, but whether it adds useful information to the standard global questions by which well-being is commonly measured.

In our view, the conceptual assumptions underlying (1) can be defended, though undoubtedly not to everyone’s satisfaction. Psychologists are more comfortable than economists when it comes to comparing indicators of feelings or utility across individuals. The facts that self-reported satisfaction is correlated with physiological measures and health outcomes and that there is some correlation between objective circumstances and affective ratings suggest that there is some

<sup>1</sup> This idea is not new to us. Greg Dow and F. Thomas Juster (1985) use this framework to analyze time-use data combined with what we call “domain-specific satisfaction” for 13 activities using the adjective “enjoy.”

signal in interpersonal comparisons of affect.<sup>2</sup> Additionally, in Kahneman et al. (2003) we find that positive and negative affect are highly correlated across situations (less so across individuals), suggesting that net affect provides an accurate characterization of situations.

### V. Conclusion

The goal of public policy is not to maximize measured GDP, so a better measure of well-being could help to inform policy. Here we propose measuring national well-being by weighting the time allocated to various activities by the subjective experiences associated with those activities. The main advantages of our bottom-up approach vis-à-vis top-down life satisfaction measures are: (i) it avoids some of the biases (e.g., duration neglect) of global retrospective evaluations; and (ii) it is connected to the allocation of time, which can be measured. If time is not allocated optimally to begin with, then well-being accounts could provide a particularly useful point of reference for society.

The NWBA can be used to summarize the average affective well-being of a population. Three potential uses are the following: (i) Changes in well-being in a country over time can be tracked, and the growth can be decomposed into a component due to changes in the allocation of time across situations, a component due to changes in affect for a given set of situations, and a residual. (ii) For subpopulations (e.g., rich vs. poor) at a given time, differences in well-being can be attributed to differences in time allocated across situations, differences in affect derived from a given set of situations, and a residual. (iii) Differences in well-being between countries can likewise be compared and decomposed.

In addition, time-based measures of well-being could also be related to individual outcomes, such as health and brain activity. Well-being accounts could help to understand how subjective experiences relate to health.

Because of adaptation and the fact that individual fixed effects (possibly genetic factors)

account for much of the variance in self-reported satisfaction, one may ask whether a NWBA index that is not particularly responsive to changes in policy or living standards is of much interest. Several responses are possible: (i) Although circumstances account for little variation in self-reported life satisfaction across subjects, the relevant consideration is how circumstances relate to the *average* level of well-being. (ii) The allocation of time changes over time, and can be influenced by policy (e.g., overtime laws); it would be useful to see how such changes map into well-being. (iii) GDP only grows by 3 percent or so each year, so small changes are typical in measures of material well-being. A large sample would be needed to detect such changes on an annual basis, however.

### REFERENCES

- Brickman, P. and Campbell, D. T.** "Hedonic Relativism and Planning the Good Society," in M. H. Appley, ed., *Adaptation-level theory*. New York: Academic Press, 1971, pp. 281–305.
- Brickman, P.; Coates, D. and Janoff-Bulman, R.** "Lottery Winners and Accident Victims: Is Happiness Relative?" *Journal of Personality and Social Psychology*, August 1978, 36(8), pp. 917–27.
- Clark, Andrew E.** "Inequality-Aversion and Income Mobility: A Direct Test." Working paper, Centre National de la Recherche Scientifique: DELTA, Paris, France, June 2003.
- Cohen, S.; Doyle, W. J.; Turner, R. B.; Alper, C. M. and Skoner, D. P.** "Emotional Style and Susceptibility to the Common Cold." *Psychosomatic Medicine*, July-August 2003, 65(4), pp. 652–57.
- Csikszentmihalyi, Mihaly.** *Flow: The psychology of optimal experience*. New York: Harper and Row, 1990.
- Davidson, Richard J.** "Affective Neuroscience and Psychophysiology: Toward a Synthesis." *Psychophysiology*, September 2003, 40(5), pp. 655–65.
- Diener, Ed.** "Subjective Well-Being: The Science of Happiness and a Proposal for a National Index." *American Psychologist*, January 2000, 55(1), pp. 34–43.

<sup>2</sup> One perhaps trivial but reassuring indication we have of the signal in interpersonal comparisons using our DRM approach is that individuals who reported sleeping less were more likely to report feeling tired during each hour of the day.

- Diener, E. and Suh, E. M.**, eds. *Culture and subjective well-being*. Cambridge, MA: MIT Press, 2000.
- Dow, Greg and Juster, F. Thomas**. "Goods, Time, and Well-Being: The Joint Dependence Problem," in F. Thomas Juster and Frank P. Stafford, eds., *Time, goods, and well-being*. Ann Arbor, MI: Institute for Social Research, University of Michigan, 1985, pp. 397–413.
- Easterlin, R. A.** "Will Raising the Income of All Increase the Happiness of All?" *Journal of Economic Behavior and Organization*, June 1995, 27(1), pp. 35–47.
- Edgeworth, F. Y.** *Mathematical psychics: An essay on the application of mathematics to the moral sciences*. London, U.K.: Kegan Paul, 1881; reprinted, New York: M. Kelly, 1967.
- Frey, Bruno and Stutzer, Alois**. "What Can Economists Learn from Happiness Research?" *Journal of Economic Literature*, June 2002, 40(2), pp. 402–35.
- Inglehart, R. and Rabier, Jacques-René**. "Aspirations Adjust to Situations—But Why Are the Dutch So Much Happier than the Germans? Subjective Well-Being in Longitudinal and Comparative Perspective," in Frank M. Andrews, ed., *Research on the quality of life*. Ann Arbor, MI: Institute of Social Research, University of Michigan, 1986, pp. 1–56.
- Inkeles, Alex**. "Industrialization, Modernization and the Quality of Life." *International Journal of Comparative Sociology*, January–April 1993, 34(1–2), pp. 1–23.
- Kahneman, Daniel**. "New Challenges to the Rationality Assumption." *Journal of Institutional and Theoretical Economics*, March 1994, 150(1), pp. 18–36.
- Kahneman, Daniel; Diener, E. and Schwarz, N.**, eds., *Well-being: Foundations of hedonic psychology*. New York: Russell Sage Foundation Press, 1999.
- Kahneman, Daniel; Krueger, Alan; Schkade, David; Schwarz, Norbert and Stone, Arthur**. "A Survey Method For Characterizing Daily Life Experience: The Day Reconstruction Method (DRM)." Mimeo, Princeton University, 2003.
- Kahneman, Daniel; Wakker, Peter P. and Sarin, Rakesh**. "Back to Bentham? Explorations of Experienced Utility." *Quarterly Journal of Economics*, May 1997, 112(2), pp. 375–405.
- Kiecolt-Glaser, J. K.; McGuire, L.; Robles, T. F. and Glaser, R.** "Psychoneuroimmunology and psychosomatic medicine: Back to the future." *Psychosomatic Medicine*, January–February 2002, 64(1), pp. 15–28.
- Lucas, Richard E.; Clark, Andrew E.; Georgellis, Yannis and Diener, Ed.** "Re-examining Adaptation and the Set Point Model of Happiness: Reactions to Changes in Marital Status." *Journal of Personality and Social Psychology*, March 2003, 84(3), pp. 527–39.
- Riis, Jason; Loewenstein, George; Baron, Jonathan; Jepson, Christopher; Fagerlin, Angela and Ubel, Peter**. "Ignorance of Hedonic Adaptation to Hemo-Dialysis: A Study Using Ecological Momentary Assessment." Mimeo, Princeton University, 2004.
- Schwarz, N. and Strack, F.** "Reports of Subjective Well-Being: Judgmental Processes and Their Methodological Implications," in D. Kahneman, E. Diener, and N. Schwarz, eds., *Well-being: The foundations of hedonic psychology*. New York: Russell Sage, 1999, pp. 61–84.
- Stone, A. A. and Shiffman, S.** "Ecological Momentary Assessment (EMA) in Behavioral Medicine." *Annals of Behavioral Medicine*, August 1994, 16(3), pp. 199–202.
- Sunstein, Cass R. and Thaler, Richard**. "Libertarian Paternalism Is Not an Oxymoron." *University of Chicago Law Review*, 2004 (forthcoming).
- Urry, H. L.; Nitschke, J. B.; Dolski, I.; Jackson, D. C.; Dalton, K. M.; Mueller, C. J.; Rosenkranz, M. A.; Ryff, C. D.; Singer, B. H. and Davidson, R. J.** "Making a Life Worth Living: Neural Correlates of Well-being." *Psychological Science*, 2004 (forthcoming).