

This PDF is a selection from a forthcoming volume from the National Bureau of Economic Research

Volume Working Title: Measuring the Subjective Well-Being of Nations:
National Accounts of Time Use and Well-Being

Volume Editor / Conference Organizer: Alan B. Krueger, editor

Volume Publisher: University of Chicago Press

Volume URL: <http://www.nber.org/books/krue08-1>

Conference Date: December 7-8, 2007

Title: Rejoinder

Author: Alan B. Krueger, Daniel Kahneman, David Schkade, Norbert Schwarz, Arthur A. Stone

Date Received: January 12, 2009

URL: <http://www.nber.org/chapters/c5061>

Rejoinder

Alan B. Krueger, Daniel Kahneman, David Schkade,
Norbert Schwarz, and Arthur A. Stone

The contributors to this volume raise several valid points about the strengths and weaknesses of our proposed method for National Time Accounting (NTA), particularly regarding the idea of measuring subjective well-being by the fraction of time people spend in an unpleasant emotional state (the U-index; see chapter 1 of this volume). To be clear, we should emphasize that in our contribution, we did not attempt to provide a comprehensive measure of all aspects of well-being. We offer a new measure of an aspect of well-being that is: (a) relevant to people's daily lives, (b) distinct and measured separately from other aspects of well-being in the existing literature, and (c) related to possible policy actions (e.g., overtime restrictions) and technological developments in society because of the link to time use.

In this brief rejoinder we concentrate on responding to the main criticisms raised. But we should not lose sight of the generally positive and encouraging reactions to the approach that we proposed, especially by J. Steven Landefeld, whose agency is charged with measuring the National Income and Product Accounts (see chapter 4 of this volume). Our goal here is to highlight what can be done to improve the measurement of evaluated time use and to clarify what our approach adds and does not add, rather than to defend our approach as the only way to proceed.

In chapter 2 of this volume, George Loewenstein states, "I believe that much if not most of what makes life worthwhile is *not* captured by moment to moment happiness, but corresponds more closely, if not perfectly, to what Krueger et al. acknowledge to be absent from NTA, namely 'people's general sense of satisfaction or fulfillment with their lives as a whole, apart from moment to moment feelings.'" This theme also emerges to a lesser extent in David Cutler's chapter (see chapter 3 of this volume). We already have acknowledged that our approach to NTA excludes one's sense of meaning

and fulfillment, although we suspect that a high sense of fulfillment will not be without positive emotional consequences. Still, we think that NTA, and the U-index in particular, capture a good deal of what makes life miserable, if not what makes it worthwhile. First, people who are in pain or depressed much of the time are probably miserable; they certainly spend their time in more restricted ways than others who are not in pain or depressed, and they express low levels of life satisfaction. Second, the approach can be extended to measure additional features of experience related to whether time use is worthwhile, such as whether people consider their specific uses of time to be a waste of time or meaningful. Third, one could perform a horse race to examine whether cumulative affective experience or self-reported life satisfaction does a better job predicting objective outcomes, such as health and mortality. We hope this test will be conducted in the future. Fourth, we think the U-index has measurement properties that are superior to standard measures of life satisfaction, such as being an ordinal measure at the level of feelings experienced in situ. Moreover, standard measures of global life satisfaction are subject to numerous contextual influences (Schwarz and Strack 1999), which are attenuated under episodic reporting conditions (Schwarz, Kahneman, and Xu 2009). Lastly, we note that even if global evaluations of life satisfaction and fulfillment are considered to provide a more accurate reflection of the extent to which life is worthwhile, experienced well-being measures still provide additional information about the emotional experience of daily life.

David G. Blanchflower raises the question of whether experienced well-being and the U-index yield many new insights beyond what has been learned from studies of life satisfaction and overall happiness (see chapter 7 of this volume). He emphasizes that results using data on either self-reported happiness or the U-index find that subjective well-being is higher for those who are older, white, married, and employed, and for those who are more highly educated and have higher income. We view findings such as these as partly validating our measure of experienced well-being. At the same time, the correlation between experienced well-being and a circumstance like household income is substantially weaker than the correlation between life satisfaction and income, suggesting that a different process relates circumstances to people's experienced happiness than to their global judgments of well-being. Indeed, the Easterlin paradox of a weak correlation between income (or changes in income) and subjective well-being (or changes in subjective well-being) seems to apply more strongly when subjective well-being is measured by experienced affect than by a judgment of life as a whole (see Stevenson and Wolfers 2008; Kahneman et al. 2006; Krueger 2008).

If the only goal of NTA was to describe people or demographic groups, then we would agree with Blanchflower that it is possible to collect subjective well-being data more efficiently than with evaluated time use. However, characterizing people is not the only goal, or even the main goal, of NTA.

An important application is to understand from where differences in well-being arise. National Time Accounting provides insight into this issue by illuminating how different sociodemographic positions are associated with different time use and different emotional experiences, providing information that is policy relevant. Moreover, cross-national comparisons on the basis of NTA data provide insight into how different organizations of daily life relate to the well-being of citizens by permitting a decomposition of differences in subjective well-being between countries into differences due to time allocation and differences due to the emotional experience of a given set of activities. For international comparisons, Blanchflower acknowledges that one obtains meaningfully different results using affect reported for episodes of the previous day—or even the previous week—and reports of overall happiness and life satisfaction. The reversal of the ranking of the French and American comparison in our chapter is a vivid example of this phenomenon, and Blanchflower provides additional data to this effect. Likewise, changes for a nation over time can be traced to changes in time use and changes in emotional experiences for a given time allocation. These decompositions are not possible with standard satisfaction data. Part of what makes life more enjoyable is spending more time in enjoyable activities; this is highlighted in NTA. We also note that none of the previous studies in the time-use literature that touched on NTA actually applied the technique to compare differences between countries or changes within countries over time.

Another goal of NTA is to characterize the emotional experience of time use during certain activities and situations. Our and others' (e.g., Csikszentmihalyi 1990; Robinson and Godbey 1997) measures of experienced well-being have added new insights in this regard. For example, we find that child care and adult care appear to be particularly unpleasant activities while they are being conducted. We also find that commuting ranks as one of the most unpleasant activities of the day, while watching television is an affectively average activity. And we find that interacting with others generally raises the emotional experience of an activity. Findings like these extend the boundaries of what has been learned from global judgments of life as a whole.

William Nordhaus maintains that there is a fundamental flaw in attempts to use subjective well-being as a social indicator (see chapter 5 of this volume). He argues that emotions, and subjective well-being more generally, are not—and cannot be—interpersonally cardinal variables. Nordhaus argues that an interpersonally cardinal variable “must have a uniquely defined zero and a well-defined unit of increment, and there must be a method to compare the values across individuals.” He further argues that the zero point (and presumably the increment) must be stable across time and people. He claims that there simply is no interpersonal scale for reporting subjective data such as happiness and pain. “Neither blue rivers nor blue moods,” he

argues, “constitute a meaningful index of emotions because they are not based on interpersonally cardinal variables.”

Before responding, it is useful to be clear about terms. Subjective data are reports of variables that only the person doing the reporting can observe. Objective data can, in principle, be observed by an external party (or parties) in addition to the person doing the reporting. Feelings are clearly subjective. No one else can experience your emotions to verify how you feel, although others can see likely correlates of your emotions (e.g., whether you smile or grimace). Life satisfaction is also an inherently subjective variable. Height, consumption, and income are objective variables. What makes objective data, like height or consumptions, interpersonally cardinal variables is not that they can be observed by a third party, however, but that a common convention is used to measure and report them. For example, height can be measured in inches or centimeters for someone in shoes or bare feet. Without the convention of a ruler, height does not meet Nordhaus’s interpersonally cardinal criteria. Even for objective variables, there are situations in which there is not an accepted convention of measurement. For example, prior to the advent of railroads and time zones in the nineteenth century, every local town set its own time; zero hour was different in different locales. Greenwich Mean Time enabled time to be measurable.

At one level, we have some sympathy for Nordhaus’ critique—indeed, the U-index was developed largely to relax some of the restrictive measurement requirements of social indicators. The U-index *does not* require a unique and universally defined zero point and increment to be a useful social indicator. It was developed precisely to avoid the need for interpersonal comparisons of interval scaled data, which is the thrust of Nordhaus’ critique. Yet at another level we disagree with his critique, even as it applies to more standard measures of subjective well-being that preceded the U-index.

Nordhaus asserts that subjective variables such as pleasure or pain and likes or dislikes are not interpersonally comparable. He asserts this on principle and provides no theoretical or empirical justification for his contention. Yet the extensive material reviewed in sections 1.3.1 to 1.3.3 of our chapter provides substantial evidence that measures of subjective experience are meaningfully related to physiological indicators and are predictive of important real-world outcomes, from marriage to immune system function to mortality. This evidence is difficult to reconcile if differences in subjective reports of well-being across subjects are meaningless because they are not interpersonally comparable. There are numerous examples where conventions of measurement have been successfully used to report and compare ratings of emotions and subjective evaluations across individuals. Consider the following scenarios. College students are routinely asked to rate the quality of their professors on a numerical scale, and the average rating across students is used for tenure and salary decisions. Netflix asks subscribers to rate how much they liked movies on a scale of one to five and then uses

this information, along with other subscribers' subjective ratings, to provide recommendations for new movies. Companies routinely survey their employees' and customers' satisfaction. Doctors in every hospital in the United States ask patients how much pain they feel on a scale of zero to ten, sometimes associating faces with the different ratings, and the responses are used to guide a course of action. Even the *Journal of Political Economy* asks referees to give a subjective rating of the quality of the paper they reviewed, from one to one hundred. Unless one believes that all of these efforts are pure folly, providing no useful information, it would seem that subjective variables pass a market test of being interpersonally cardinal. In sum, making interpersonal comparisons of individuals' subjective ratings has proved a valuable and enduring practice in numerous fields, and the mere fact that the cardinality criteria that Nordhaus lays out are hard to substantiate does not imply that the measures fail to capture meaningful information.

At a conceptual level, thousands of years of evolution have probably abetted the development of conventions to enable people to communicate and convey the intensity of their emotions. It is in one's survival interest to be able to detect and express how much something hurts, for example. The socialization process also guides people to express the strength of their emotions in an understandable way. Verbal descriptions of feelings come to have somewhat common meanings, although there can be a lot of noise in the way people express themselves. Nonetheless, this process enables interpersonal measurement conventions to be established for subjective variables. It is also worth noting that in surveys, it is common to give respondents verbal anchors to guide them (e.g., a zero means the feeling was not present, and a six means it was very much part of the experience) so they have a common zero point and a sense of what the interval between scales is in reporting subjective responses.¹ Although we would not push this argument too far, there are reasons to believe that social conventions can make it possible to report and contrast emotions.

We recognize, however, that language and custom can affect the convention that is used to report subjective variables. Different societies develop different conventions. Indeed, we argue in chapter 1 that this is an issue for comparisons of life satisfaction between France and the United States. This is one reason why we proposed the U-index. The U-index is robust to the interpersonal measurement convention, as long as a given person uses the same convention for positive and negative emotions.

As Nordhaus acknowledges, the U-index "would appear to avoid the difficulties of some happiness indices by its creation of an ordinal index."

1. It seems to us that the absence of feeling an emotion like pain *does* provide a natural zero point, even if the width of the interval of increments may be vague. Thus, we think it should be noncontroversial to develop an index that measures the percentage of time that people spend in some pain.

A measurement parable

One of the anonymous reviewers of this volume suggested the following response to William Nordhaus's claim that "hedonic measures do not meet the standards for an interpersonally cardinal variable that are required to construct a meaningful quantitative social indicator."

Imagine a world where lots of people smoke. However, this world has not progressed enough scientifically to have anything like twenty-first century measures of health. All that this world has achieved, sad to say, is some rough subjective measures of health. There are in this world some surveys that look at just those. In them, human beings fill out forms where they report how they feel in response to questions such as "My health is excellent . . . fairly good . . . poor . . . very poor?" and they give other social and economic data. But there are no blood test readings or heartbeat count or scans or anything like that. . . . But they can, in this world, run regression equations. Their dictator must have been a theoretical econometrician.

A commentator of the day, called BN, makes a big speech and says there is no point in trying to use these subjective health measures for anything. You should all pack up and go home, he says.

But, to show he is wrong, a group of researchers tries to estimate Subjective Health equations and they find that smoking comes in with a big negative coefficient, whether controlling for everything else or not controlling (it does, incidentally, if you estimate Subjective Health equations on twenty-first century data). They then prescribe anti-smoking restrictions. BN writes complaining letters, lamenting the end of the scientific measurable method, to the *New York Times*, but the researchers press ahead. Millions of lives are saved. They become heroes. Yet according to BN not a single interpersonally cardinal health indicator exists in this world.

The referee's parable, which is not so far removed from reality, highlights the point that progress has been made by comparing individuals' subjective evaluations of their health and other domains of life.

*In his original draft, Nordhaus used the phrase "measurable variable" instead of "interpersonally cardinal variable." We have edited the referee's passage to accord with the revised version of Nordhaus's paper.

However, he argues that our procedure "simply pushes the difficulty into the background." To us, the appeal of the U-index is that different people do not have to use the same convention to measure their emotions, as long as the emotion that they rate highest is the one that they feel most intensively at the time. Stated simply, the requirement for the U-index is for someone to be able to decide at a given moment if they are feeling more happy than sad

or more pain than pleasure. Nordhaus argues that the intensity of emotions cannot be compared because there is no conceivable zero point or increment for emotions, even for a given person at a given moment in time.

Nordhaus accepts that emotions can satisfy an ordinal ranking, presumably meaning that someone can determine that he or she feels more or less pain in a given situation. He does not believe that it is possible for someone to decide whether he or she feels more pain than pleasure during that situation, however. Thus, the runner who reports in our surveys that while jogging, his pain is high but his happiness is even higher is not providing meaningful information, according to Nordhaus; nor is the runner who says he felt more pain than pleasure when he sprained his ankle. It is not clear, however, why ordinality would apply within emotions but not between them. Emotions have some properties in common. If the human brain is capable of deciding that something hurts more or less in a given situation, why can it not decide that a given situation is more painful than pleasurable?

No evidence is presented to substantiate Nordhaus' claim that the strength of emotions at a point in time cannot be compared, or that in principle, there is unlikely to be a natural zero point for pain and other emotions. Indeed, Nordhaus implies that no evidence (such as the correlation between self-reported emotions and brain imaging) could persuade him that emotions can be compared, because they are not measurable variables. His argument rests on the presumption that the (conceptual) zero point and increment for measuring emotions "will vary with mood, circumstances, genetics, context, history, and culture." This is a more difficult argument to defend when it comes to the U-index, however, as the U-index tries to measure mood as an outcome, and the zero point and increment can be person specific for the U-index—so genetics, history, and culture are not stumbling blocks. While the factors that Nordhaus raises may well add noise to the measurement of the U-index, they do not seem to make it meaningless for individuals to rate the intensity of how they feel at a point in time along various affective dimensions.

To the extent that one considers evidence relevant, the evidence does suggest to us that there is much useful signal in the U-index, and evaluated time use more generally. As detailed in our chapter, reports of the intensity of emotions across individuals do correlate with physiological measures. If self-rated emotions were not comparable across people, at least to some extent, we would expect a correlation of zero. In addition, the pattern of the U-index across demographic groups and activities is, for the most part, intuitive. Finally, cognitive interviews indicated that subjects selected the affective dimension that they assigned the highest numerical rating to as the most intense feeling they had during the episode.

Nordhaus misrepresents the U-index when he writes, "This approach is equivalent to assuming that there are interpersonally cardinal subindices in an underlying preference function, $U(P, H)$." The underlying preference

function does not need to be interpersonally cardinal—it can vary across individuals. Moreover, $U(P, H)$ can vary for a given individual over time, and it is unnecessary for the researcher to specify the $U(P, H)$ function for the U-index to be a meaningful social indicator. The experience of a given person feeling more pain than pleasure is of relevance even if the underlying preference function changes.

Nordhaus goes beyond the requirements for interpersonal cardinality in criticizing the U-index, because the U-index is an ordinal measure at the level of emotions. He argues that “blue moods” or unpleasant experiences cannot conceivably be defined or measured because there is no natural zero point or standard increment for a given person’s emotions.² In this view, no latent variable can conceivably indicate a person’s likes and dislikes or pleasure and pain. However, this standard would seem inconsistent with the underpinnings of the “standard ordinal preference function” as well. If a person can decide that one bundle is preferred to another, then Nordhaus would presumably accept that there is an underlying latent variable with a common zero point and well-defined increment that enables the two bundles to be compared. Thus, the extent to which someone liked something would have to be a conceivable latent variable for that person to decide that he or she preferred one bundle over another and therefore chose it (presuming that people choose the bundle they like most). The only difference in measurement requirements between the U-index and the standard ordinal preference function is that in the former, a person is assumed capable of comparing whether he or she is more happy than sad at a given time, and in the latter, the person is assumed capable of comparing how much he or she would like alternative consumption bundles that he or she may or may not consume.

All social indicators require assumptions and entail some noise and uncertainty. The assumptions underlying our proposal for National Time Accounting seem to us to strike a reasonable balance between measurement requirements and practicality. We did not develop the U-index from first principles as a comprehensive indicator of the well-being of society. Instead, we offer it as a plausible indicator of the relative frequency of misery experienced in certain settings and by various groups. We hope that the U-index and related indicators can provide a useful indicator of situations that are associated with unpleasant emotional experiences and of groups that are more likely to endure emotionally unpleasant experiences. We would not expect a goal of public policy to be to minimize the U-index, but instead for the U-index to highlight areas that are worth further investigation. We also hope that NTA can provide a means for tracking

2. Presumably, to define a blue mood, all one would need is a zero point—the absence of feeling blue—because the width of the increment is irrelevant if the goal is to derive an indicator of the presence or absence of any nonzero level of the emotion.

whether societies are spending their time in more or less enjoyable ways, which can be an input along with others to derive a picture of the progress of society.

In conclusion, it is useful to recall Jan Tinbergen's (1976) advice: "Progress in our understanding can only be based on the push for measurement of phenomena previously thought to be non-measurable" (51).

References

- Csikszentmihalyi, M. 1990. *Flow: The psychology of optimal experience*. New York: Harper Collins.
- Kahneman, D., A. Krueger, D. Schkade, N. Schwarz, and A. Stone. 2006. Would you be happier if you were richer? A focusing illusion. *Science* 312 (5782): 1908–10.
- Krueger, A. 2008. Comment on Stevenson and Wolfers, "Economic growth and subjective well-being: Reassessing the Easterlin paradox." *Brookings Papers on Economic Activity* (Spring): 95–100. Washington, D.C.: Brookings Institution.
- Robinson, J., and G. Godbey. 1997. *Time for life: The surprising ways Americans use their time*. University Park Pennsylvania State University Press.
- Schwarz, N., D. Kahneman, and J. Xu. 2009. Global and episodic reports of hedonic experience. In *Using calendar and diary methods in life events research*, ed. R. Belli, D. Alwin, and F. Stafford, 157–74. Newbury Park, CA: SAGE.
- Schwarz, N., and F. Strack. 1999. Reports of subjective well-being: Judgmental processes and their methodological implications. In *Well-being: The foundations of hedonic psychology*, ed. D. Kahneman, E. Diener, and N. Schwarz, 61–84. New York: Russell-Sage.
- Stevenson, B., and J. Wolfers. 2008. Economic growth and subjective well-being: Reassessing the Easterlin paradox. *Brookings Papers on Economic Activity* (Spring): 1–87. Washington, D.C.: Brookings Institution.
- Tinbergen, J. 1976. More empirical research. In *Economics in the future*, ed. K. Dopfer, 39–52. London: Macmillan Press.