

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

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

Volume Author/Editor: Alan B. Krueger, editor

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-45456-8

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

Conference Date: December 7-8, 2007

Publication Date: October 2009

Title: That Which Makes Life Worthwhile

Author: George Loewenstein

URL: <http://www.nber.org/chapters/c5054>

That Which Makes Life Worthwhile

George Loewenstein

Too much and for too long, we seemed to have surrendered personal excellence and community values in the mere accumulation of material things. Our Gross National Product, now, is over \$800 billion dollars a year, but that Gross National Product—if we judge the United States of America by that—that Gross National Product counts air pollution and cigarette advertising, and ambulances to clear our highways of carnage. It counts special locks for our doors and the jails for the people who break them. It counts the destruction of the redwood and the loss of our natural wonder in chaotic sprawl. . . . And the television programs which glorify violence in order to sell toys to our children. Yet the gross national product does not allow for the health of our children, the quality of their education or the joy of their play. It does not include the beauty of our poetry or the strength of our marriages, the intelligence of our public debate or the integrity of our public officials. It measures neither our wit nor our courage, neither our wisdom nor our learning, neither our compassion nor our devotion to our country, it measures everything in short, except that which makes life worthwhile. —Robert Kennedy, 1968 (quoted from Krueger et al., Chapter 1, this volume)

2.1 Introduction

Like Krueger and his collaborators, I find Robert Kennedy's words both compelling and moving. I share with Kennedy and Krueger et al., the view that gross national product (GNP) fails to fully capture that which makes life worthwhile, and Kennedy's list of what GNP fails to include comes close

George Loewenstein is the Herbert A. Simon Professor of Economics and Psychology at Carnegie Mellon University.

I thank Erik Angner and Alan Krueger for helpful comments.

to matching my own perspective on what makes life worthwhile. The issue I address in this commentary is the degree to which National Time Accounting (NTA) captures what makes life worthwhile, including, given the prominence they give to Kennedy's quote, the aspects listed by Kennedy.

The purpose of a scale, or an index such as NTA's U-index, is to measure an underlying construct. For the U-index, the underlying construct is welfare. As Krueger et al. express it, NTA provides "an alternative way of measuring society's well-being, based on time use and affective (emotional) experience." Ideally, the U-index could be used to determine whether one group of people (e.g., the citizens of a country) is better off than another, or whether a specific group is, or would be, better off under one set of circumstances than another.

In the language of research methodology, Kennedy's passage can be interpreted as a critique of GNP's *validity* as an index of welfare. Validity addresses the degree to which an index or scale measures the construct that it is intended to measure. (Reliability, in contrast, addresses the extent to which you get the same answer when you elicit the scale in different ways or at different points in time.)

Scales and indexes have low validity to the extent that they encompass dimensions that are not part of the construct they are intended to represent, and fail to encompass dimensions that are part of the construct. As summarized in table 2.1, Kennedy can be interpreted as having made the point that GNP has low validity as a measure of welfare because it includes a variety of things that do not belong in the construct of welfare (top right cell), and fails to encompass many important factors that are important aspects of welfare (bottom left cell).¹

Beyond the specifics of what it should include that it does not, and what it does not include that it should, GNP embodies implicit assumptions that are questionable. For GNP to represent a reasonable proxy for welfare, the economic activity indexed by GNP must be allocated to purposes that people value. If people or their elected governments do a poor job of allocating wealth to activities that enhance their well-being, by whatever metric of well-being one adopts, then GNP will fall short of measuring welfare.

1. Kennedy does not mention material prosperity, but presumably he would agree that prosperity is a part of GNP and deserves to be considered as one dimension of welfare. Hence, I have included it in the on-diagonal (bottom right) cell of the table. I leave the top left cell blank because there are an infinite number of things that do not belong in GNP that are not included in it, such as the length of people's hair. Krueger et al. expand on Kennedy's list of things that are not accounted for (or accounted for improperly) in GNP, citing "near-market" activities (e.g., unpaid cleaning, cooking, and child care), social activities, consumer surplus (because economic activity is measured by prices, which reflect marginal valuations), prices distorted by imperfectly competitive markets, the distribution of income (which might matter in its own right and might also influence prices and marginal valuations in a fashion that could distort welfare calculations), and finally externalities (costs people impose on others that they do not internalize).

Table 2.1 What makes life worthwhile: Kennedy versus National Income Accounts

Kennedy	GNP	
	Not included	Included
Not included	(Infinite)	<ul style="list-style-type: none"> • Air pollution • Cigarette advertising • Ambulances to clear highway carnage • Special locks for doors • Jails • Destruction of redwood • Chaotic sprawl • Violent television programs
Included	<ul style="list-style-type: none"> • Health of children • Quality of their education • Joy of their play • Beauty of our poetry • Strength of marriages • Intelligence of public debate • Integrity of public officials • Wit, courage, wisdom, and learning • Compassion • Devotion to country 	Material prosperity?

That is, indexing welfare by GNP assumes, implicitly, that people allocate resources in a fashion that promotes their welfare.²

2.1.1 Validity of NTA and the U-index

What about NTA? Is NTA and its instantiation in the U-index more successful than GNP in capturing what makes life worthwhile? National Time Accounting does have desirable qualities. Most fundamentally, NTA, unlike GNP, does not assume that people necessarily behave in a self-interest fashion. In the not-so-old (and definitely not-so-good-old) days of economics, when it was widely assumed that people were reliable pursuers of self-interest, measuring welfare was (comparatively) easy. Ignoring distributional issues, it could be assumed that increasing disposable income also increased well-being because it presumably gave people greater scope to pursue their own material and nonmaterial goals. The emergence of behavioral economics, with its multiple challenges to the view that people rationally pursue self-interest, complicated this tidy picture. Once one accepts that people are unreliable, and indeed often biased, pursuers of self-interest, it can no longer be assumed that increasing affluence will make them better off.

2. See Loewenstein and Ubel (2008) for a deeper discussion of this point.

As the economist Avner Offer (2006) points out in a recent book titled *The Challenge of Affluence*, if people make systematic mistakes when it comes to maximizing their own well-being, then increasing their income may not only not enhance their welfare; it may be tantamount to giving them more rope to hang themselves with. Offer cites research on well-being (albeit not using NTA methods) that he interprets as showing that well-being decreased in the United States and Britain not only as, but *because*, affluence increased. Perhaps NTA's greatest strength, then, is that it does not assume any particular relationship between income and happiness. Has the increase in income over the last century led to improvements in welfare? National Time Accounting would indicate that it has only if people spend a larger fraction of their time in a predominantly positive mood.

However, NTA and the U-index do have serious limitations. Krueger et al. acknowledge that "Like the National Income Accounts, NTA is also incomplete, providing a partial measure of society's well-being." NTA, they note, "misses people's general sense of satisfaction or fulfillment with their lives as a whole, apart from moment to moment feelings." Nevertheless, they argue, NTA "provides a valuable indicator of society's well-being, and the fact that our measure is connected to time allocation has analytical and policy advantages that are not available from other measures of subjective well-being, such as overall life satisfaction." As depicted in figure 2.1, Krueger et al.'s implicit perspective seems to be that, while NTA, like NIA (national income accounts) misses some important aspects of welfare, it is superior to NIA in terms of capturing "true" welfare.

My own perspective is somewhat more pessimistic. Contrary to figure 2.1, and more consistent with figure 2.2, 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." In the remainder of this chapter, I provide a more detailed rationale for my misgivings about NTA, starting with the next section, which enumerates dimensions of welfare that are missing from NTA.

Table 2.2 summarizes Kennedy's perspective on what makes life worthwhile, classifying his specific items into broad categories. The two main categories that subsume the majority of his items are *wisdom* (with four items) and *values* (encompassing four or five items, depending on whether the "strength of our marriages" falls under this heading). Only one item ("the joy of [our children's] play"—in italics) is directly related to happiness.

It is possible that four other items (highlighted in bold) could be reflected in happiness in an indirect fashion. That is, it seems reasonable to assume that children and their parents are happier when children are healthy. Perhaps more controversially, it might be expected that people would be happier when marriages are stronger (although not if strong marriages means that

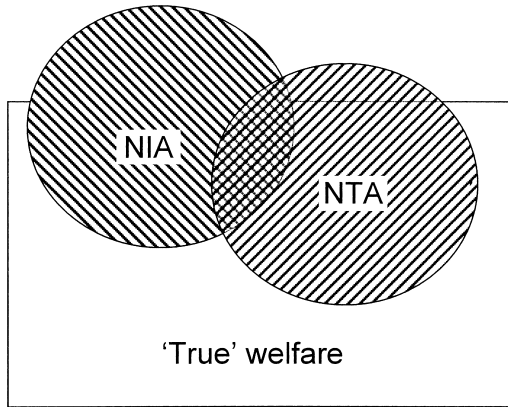


Fig. 2.1 Krueger et al.'s (implicit) perspective

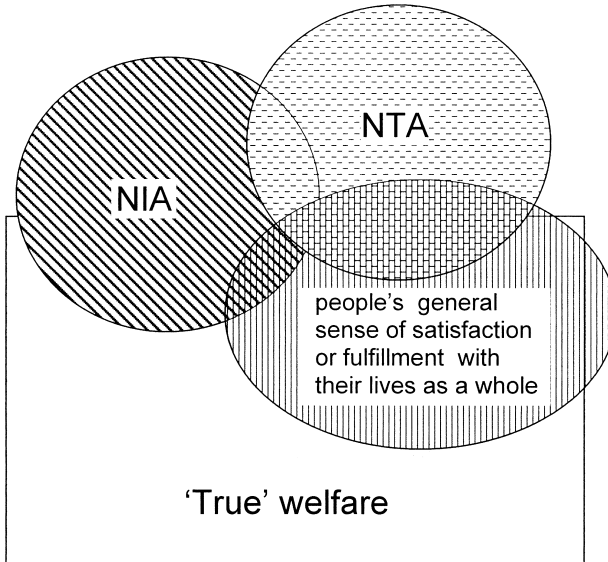


Fig. 2.2 A more pessimistic view

people are trapped in unhappy marriages), or when they are exposed to others' compassion or wit (although gratitude is not always such a pleasant feeling, and too much wit coming from others can be depressing for those lacking in it).

Even if health, social stability (a generalization of marital strength), wit, and compassion do have a positive impact on happiness, however, it seems unlikely that this impact adequately captures their full value. Thus, as I dis-

Table 2.2 Kennedy's view of what makes life worthwhile

	Health	Wisdom	Happiness	Culture	Values
Health of our children^b	X				
Quality of their education ^c		X			
<i>Joy of their play^a</i>			X		
Beauty of our poetry ^c				X	
Strength of our marriages^b					?
Intelligence of our public debate ^c		X			
Integrity of our public officials ^c					X
Wit^b		X			
Courage ^c					X
Wisdom and learning ^c		X			
Compassion^b					X
Devotion to our country ^c					X

^aDirectly captured by NTA.

^bCould be captured by NTA (indirectly).

^cUnlikely to be captured by NTA.

cuss later, although people adapt to health problems as severe as quadriplegia and exhibit close to normal levels of happiness, most people, including quadriplegics themselves would be willing to make tremendous sacrifices to retain (or regain) the use of their limbs. This suggests that unhappiness does not capture the full (negative) value of quadriplegia. It is also *possible* that an improvement on any or all of the remaining seven items (those not italicized or highlighted in bold) might increase happiness, but the connection seems tenuous at best. Are societies that are more courageous happier? Perhaps, but on the face of it this seems no more likely than the opposite.

In sum, NTA does capture some aspects of welfare that are not part of NIA, but whether it constitutes an improvement or even that much of a useful complement, is unclear.

2.2 What's Missing from NTA?

One modern perspective on what can go wrong in survey design applies insights from research on conversational norms (e.g., Grice 1975; Clark and Clark 1977) to understanding how survey respondents make sense of the questions they are asked (Clark and Schober 1992; Schwarz 1999). According to this perspective, a survey can be viewed as a kind of "conversation" between the surveyor and the respondent in which the usual norms of conversation apply.

As an illustration, the "maxim of quantity" (Grice 1975), which enjoins speakers to provide information that is new and not redundant, can shed light on the results of a study on marital satisfaction and life-satisfaction (Schwarz, Strack, and Mai 1991). Some respondents to a survey were first

asked how satisfied they were with their life as a whole and then were asked to report their satisfaction with their marriage. Others were asked the same questions in the reverse order, and still others had both questions introduced by a join lead-in designed to evoke the norm of nonredundancy by informing respondents that they would be answering two questions, one related to well-being and the other relating to their marriage. When the life satisfaction question was asked before the marital satisfaction question, the correlation between the two items was lower ($r = .32$) than when they were asked in the reverse order ($r = .67$), presumably the marital question brought the marital dimension of life to mind when people were reporting their overall life satisfaction. However, the correlation was lowest ($r = .18$, n.s.) in the third condition, presumably because, as Schwarz (1999) expressed it, “respondents interpreted the general life-satisfaction question as if it were worded, ‘Aside from your marriage, which you already told us about, how satisfied are you with other aspects of your life?’”

Playing on the idea of a survey as a “conversation,” I propose a new notion of validity: validity as *feeling understood* by a researcher—a concept that, I believe, has not previously been suggested in the literature. The criterion of feeling understood can be viewed as a high-level inference, on the part of the respondent, that the “conversation” with the surveyor has the capacity to answer the surveyor’s question in a fashion that the respondent deems reasonable.

In the course of life I have been asked to complete myriad surveys that, it was apparent, were intended to measure a wide range of things: How happy was I with a class I had taken? Was I satisfied with my new car? Was giving blood a pleasant or unpleasant experience? Deducing the purpose of a scale from the questions I am being asked, I often find myself thinking that the designers of the scale have asked the wrong questions to address whatever they seemed to be interested in. I was once asked, for example, whether the telephone operator at a U.S. Airline was courteous, and my answer was affirmative; yet I suspected that the airline would have also been interested in whether I found the operator competent—whether he had been able to do for me what I needed done—and I knew that the answer to this question would have been much less favorable. The airline survey, therefore, would have fared poorly on the “feeling understood” measure of validity.

I introduce this new, and perhaps somewhat atheoretical notion of validity because, at different points in time after I had taken on the assignment of writing this commentary, I attempted to assess whether the U-index would successfully capture my own perception of the quality of a particular activity I was engaged in—that is, whether researchers who attempted to elicit my U-index using the Day Reconstruction Method (DRM) would come to the same conclusion as I would have about my quality of life in that period of time.

If there was a discrepancy, of course, it is possible that the U-index cor-

rectly assessed my well-being while I misestimated it. But this possibility actually goes to the heart of my main misgiving about the U-index. The U-index assumes that the quality of a person's life can be measured in terms of happiness, but individuals might have very different criteria for what makes their own life worthwhile. If an individual values something other than happiness, who is to say that happiness is the right measure of welfare? In the remainder of this section, I discuss a variety of dimensions of life other than happiness that I personally care about but that would not be fully picked up on by a measurement of happiness.

2.2.1 Meaning

If you asked men of my father's generation to relate their life story, a typical narrative would devote hours to the individual's experiences during the war, then devote little more than a sentence to the remaining bulk of their lives—for example, "When I got back from the war, I finished school, got married, had kids, retired, and here I am." Their experiences during the war may not have been pleasant, but they gave their lives meaning. My own father's case was especially extreme. He spent part of the war interned in a French prisoner of war camp, hungry to the point where he dug up worms for food and chewed on shoe leather. But he once reported to me that being in the camp was the peak experience of his life.

Of course memory has a way of blotting out the misery—the hunger, discomfort, and fear—and leaving an idealized residue of meaning. Yet it would be a mistake to entirely dismiss these retrospective evaluations. For my father, having to use his wits to survive in the camp and the feeling of camaraderie and interdependence with the small group he allied himself with were never matched by the comfortable suburban existence he eventually established for himself and his family. Not only is the U-index unlikely to pick up on the value from experiences such as war (or mountaineering³); it would be likely to encode as maximally negative many of the experiences that people recount as having been the most worthwhile because, while often difficult at the time, they conferred meaning.

Meaning can, of course, have many interpretations (see Karlsson, Loewenstein, and McCafferty [2004] for a discussion of the nonrole of meaning in economics⁴). However, many of the possible interpretations of the concept

3. The reports of mountaineers are similar. When they give a more complete response to the "why?" question than "because it is there," mountaineers often cite meaning as a major benefit they derive from their escapades (Loewenstein 1999).

4. We distinguish between four different possible definitions, which we discuss in order from the one that is easiest to assimilate with traditional economic theory to the one that is most difficult to assimilate:

- Meaning as a resolution of uncertainty about preferences: People are often uncertain about what they want from life. Finding meaning, in some cases, can entail learning about what one values or cares about.

are missing from NTA. For example, it could be argued that meaning entails having a range of emotional experiences; always being in an unchanging emotional state would entail a lack of meaning. Or, it could be argued that meaning arises from experiences that change one's self-concept or alter the story one would tell about one's life. National Time Accounting, and especially its instantiation in the U-index, which implicitly defines welfare as the absence of negative emotions, misses out on all of these notions of meaning, and especially interpretations that are associated with emotional range.

Tibor Scitovsky, one of a small number of economists who embraced psychology in the 1970s, would probably not have been a fan of this aspect of the U-index. In his classic, *The Joyless Economy* (1976), Scitovsky argued, much as do Krueger et al., that GNP is a poor measure of a society's welfare because societies often spend resources in ways that are not conducive to true well-being. However, he cautioned against the tendency for individuals and societies to expend their resources on things that bring bland "comfort," characterized mainly by an absence of risk, discomfort, or uncertainty, as compared with goods and activities that bring "pleasures," which he defined in terms of features such as challenge, risk, and variability. The U-index, which encodes only periods of net negative affect, and fails to give credit for the more dramatic ups and downs that give life much of its richness, would evaluate favorable exactly the kind of society and lifestyle that Scitovsky cautioned against.

Another likely skeptic of the U-index would have been Aldous Huxley, whose classic novel *Brave New World* presented a vision of a future dystopia in which everyone was happy because society has been engineered (partly with the aid of a drug called Soma, eerily similar to modern antidepressants) to eliminate negative emotions. As a world leader going by the title of "The Controller" states, presumably referring to a historical period coming shortly after our own, "Our ancestors were so stupid and short-sighted that when the first reformers came along and offered to deliver them from those horrible emotions, they wouldn't have anything to do with them" (45).

-
- *Meaning as an extension of self either socially or temporally:* One's life can often seem insignificant and inconsequential when viewed in the context of the span of human (or even natural) history or of the vast numbers of people alive in the world. The quest for higher meaning may serve the function of expanding the self through time and across persons.
 - *Meaning as an act of sense-making:* The brain is a sense-making organ, and one of its most important tasks is to make sense of the life of its owner. Such sense-making typically takes the form of a narrative—a "life story."
 - *Meaning as an assertion of free will:* People derive personal meaning from the act of making autonomic choices. Hence, meaning-making can involve the assertion of free will.

To this list, perhaps should be added "*meaning as the experience of a range of emotions.*" Part of what it means to be alive is to experience a range of emotions. Such a desire to experience a range of emotions may help to explain why we voluntarily expose ourselves to emotions that are normally seen as "negative" (such as the fear of a roller coaster or the sadness of a tragedy).

Huxley's implicit, although not too subtle point is that "those horrible emotions" actually have value; they are what protect us from an existence devoid of meaning. In one representative scene in the book, the two-person flying machine occupied by Henry and Lenina (both prototypical citizens of their time) suddenly rises, buoyed by a column of hot air from the chimney of a crematorium they have passed over. Huxley writes that "Henry's tone was almost, for a moment, melancholy. 'Do you know what that switchback was?' he said. 'It was some human being finally and definitely disappearing.' He sighed. Then, in a resolutely cheerful voice, 'Anyhow,' he concluded, 'there's one thing we can be certain of; whoever he may have been, he was happy when he was alive. Everybody's happy now'" (75). Huxley's *Brave New World* would achieve an almost perfect score on the U-index, despite his own intention to present it as the antithesis of true welfare.

2.2.2 Wisdom

One of the most common critiques of happiness as a measure of welfare involves the tension that often seems to exist between happiness and intelligence or wisdom. Most famously, John Stuart Mill, while embracing utilitarianism and its central assumption that happiness should be the goal of public policy, argued that the *quality* of happiness has to be considered as well as the quantity. According to Mill, although a pig might derive a great quantity of pleasure from wallowing in the mud, "it is better to be a human being dissatisfied than a pig satisfied; better to be Socrates dissatisfied than a fool satisfied" (Mill 1871, chapter 2). National Time Accounting accounts for wisdom in a positive fashion only to the degree that wisdom contributes to happiness (or, more precisely, subtracts from unhappiness), but, as suggested by the commonplace that "ignorance is bliss," a wiser society might well be a less happy one.

According to the empirical analysis presented in Krueger et al., education is among the least enjoyed activities measured by the U-index. It is the second-to-worst activity without controlling for individual fixed effects and third-to-worst after controlling for fixed effects. Yet people seem to value education tremendously. Education is a voluntary activity and is heavily subsidized by the state.

People do obtain education in part to secure professional goals, and societies certainly value education in part for economic reasons. Gross national product would be an appropriate index for capturing the economic value of education. People may also obtain education, in part, because they believe it will bring happiness in the long run. Indeed, there is suggestive evidence from one of the empirical studies presented by Krueger et al. (table 1.8 from their chapter) that this might be the case. Those with a college degree or greater have substantially lower U-indexes. Just as traditional income accounts can be a useful way of picking up on intertemporal tradeoffs of

income, NTA could be a useful tool for picking up on these intertemporal tradeoffs of happiness.⁵

However, even if education makes people miserable while they are engaged in it, people seem to value education, or the wisdom it confers, for other than either purely economic or purely hedonic reasons. My own university, like many others, offers a whole program of education targeted at senior citizens that is so popular that it has a wait-list half as long as the number of active participants. Why are so many people who have little to gain in terms of either future economic returns or happiness engaged in so much education if it leads to so much negative affect? Like meaning, wisdom seems to be a quality that people value in themselves and others, regardless of its impact on happiness.

Wisdom adds an important dimension to life. Much as gaining sight for a blind person would allow the individual to perceive dimensions that he or she had not previously perceived, even if it did not enhance their happiness, wisdom adds dimensions to thought and perception. Thus, an individual who, by dint of education, gains a taste for and appreciation of the subtle differences between wines may end up enjoying the average bottle of wine less. However, gaining a taste in wine is like speaking a new language. Dimensions of wine that were not previously apparent come into focus, and perceiving these dimensions has value in its own right. I would argue that the same is true for most forms of wisdom.

Krueger et al. were certainly acutely aware of the problem posed by Mill's objection to Bentham's utilitarianism, which may be why they included "interest" as one of only two positive affects in the short list of six affects that they measured (with pain, happy, tired, stressed, and sad being the others), even though "interest" is rarely treated as an affect by emotion researchers, seems difficult to compare to the other affects, and is not even necessarily positive. For example, I might be very *interested* to hear the details of a referee report that my coauthors have informed me is negative, yet not derive much pleasure from that interest or from the information when I obtain it. Wisdom is an important component of what makes life worthwhile, but including "interest" in the list of affects is unlikely to value wisdom appropriately.

2.2.3 Values

My wife and I spent last Thanksgiving vacation with her family, in Florida, with much of the family's time devoted to taking care of her ailing father. If

5. Of course, it is unclear which way the causality runs, or whether the benefits of higher education might come through income, which would be captured by traditional national income accounting. Without conducting extended longitudinal research, and without randomly assigning people to get different levels of education (which is probably impossible), these issues are unlikely to get resolved.

the quality of our days during that vacation had been elicited using DRM or PARS, our vacation would have come out very unfavorably. Much of the caregiving elicited strong negative emotions, whether it was because of the specifics of what the care involved or because of the contrast we were forced to confront between her father's current condition and his past vitality. Yet our low U-index during the vacation would fare badly on the *feeling understood* criterion of validity. By caring for their aging parent, my wife and her siblings were displaying their humanity, sharing their love for their father and their sense of the family as an integral unit. None of these values would have been picked up by ratings of momentary happiness.

In one of the empirical studies reported in Krueger et al., the single activity that comes out worst on the U-index, whether or not one controls for fixed effects, is adult care. Does this mean that we could improve welfare by spending less time taking care of our parents? We would, of course, need to take account of the happiness of the people being taken care of, but it seems unlikely that their welfare gains compensate for the losses of those doing the caretaking.⁶ Indeed, for many of those receiving care, it is difficult to discern if they are even aware of the fact that they are being taken care of. Should we dismiss caretaking of other people if the U-index fails to show commensurate benefits to those being taken care of? Clearly, this would be a mistake. I can easily imagine Kennedy having included "care for our elderly" in his list of what makes life worthwhile, but assuming it comes from family members instead of professionals, neither GNP nor NTA value it positively.

2.2.4 Capabilities

In a recent study, my coauthors and I (Smith, Loewenstein, and Ubel, forthcoming) asked seventy-one patients who had received a colostomy (an operation in which the bowels are surgically diverted to empty into a bag) to report two measures of happiness (a five-item satisfaction with life scale (Diener et al. 1985), and a "ladder scale" (Cantril 1967) at three points in time: (a) one week after they were released from the hospital; (b) one month after release, and (c) six months after release. The critical variable of interest was whether the colostomy was of a type that is permanent (can never be reversed) or was potentially reversible at some point in the future. Based on prior, albeit more anecdotal, evidence, we anticipated that those who had reversible colostomies would fare worse happiness-wise than those who had irreversible ones. As shown in figure 2.3, our prediction was strongly supported. Those with permanent colostomies got progressively happier over time. Those with reversible ones got less happy according to

6. Another, I believe, implausible account of why we take care of parents is to set a good example for children in the hope they will take care of us and improve our U-index when we ourselves age. It would be easy to examine whether children who do not themselves have children are less dutiful caretakers of their parents. I doubt this is the case.



Fig. 2.3 Happiness over time: Permanent versus reversible colostomy

one measure, or remained at a roughly constant level of happiness according to the other.

Should physicians react to these results (assuming they were confirmed by additional research) by ceasing to perform potentially reversible colostomies? Obviously not. Not having a colostomy is better than having a colostomy—much better—even if those with permanent colostomies are no less happy.

The improvement in happiness of the permanent colostomy group is emblematic of a large body of research showing that people adapt to a wide range of conditions—including conditions that most people would classify as extremely adverse—and come to achieve close-to-normal levels of happiness. Yet, as Peter Ubel and I discuss in a paper devoted to the point (Loewenstein and Ubel 2008), there is widespread agreement, not only by the general public, but also by people who currently have these health conditions as well as people who had them in the past, that these health conditions are extremely undesirable—a distaste that is reflected in all of these groups' stated willingness to make various types of sacrifices (e.g., willingness to pay money or to risk a chance of death) to maintain or regain health. The research showing that people powerfully dislike health conditions that they fully adapt to poses a serious challenge to measures of welfare based on happiness.

Amartya Sen (1985, 1992) and Martha Nussbaum (2000) have proposed an approach to measuring welfare that is designed to avoid exactly this problem. Their "capabilities" approach was designed to deal with the problem that people may adapt to, and hence be content with, poor social and physical conditions or injustice, because they have experienced them for a

prolonged period of time or have never experienced anything else. As Nussbaum (2000, 114) expresses it, aspirations for a better life can be squelched by “habit, fear, low expectations, and unjust background conditions that deform people’s choices and even their wishes for their own lives.” The capabilities approach delineates a series of central human capabilities, such as health, freedom from assault, political voice, property rights, equal employment, and access to education that are seen as central to welfare regardless of their connection to happiness. Several of these capabilities would be likely to be undermined by disability; hence the capabilities approach would view adverse health conditions as negative outcomes, even if those experiencing them displayed normal levels of happiness.

It is interesting to note that very few, if any, of the quality of life indicators used to rate, for example, the best city to live in or to visit, measure time use. Instead, consistent with the capabilities perspective, they tend to involve some kind of crude weighting of desiderata such as income, health, freedom, political stability, absence of crime, education, opportunities for advancement, culture, and so on.⁷ And although different quality of life measures use somewhat different criteria, they tend to produce fairly similar rankings that, at least to my eye, often seem quite reasonable, given my experience with cities.

2.3 Other Problems with NTA

2.3.1 Does National Time Accounting Get the Accounting Right?

One of the most important empirical investigations of the U-index presented in Krueger et al. is a comparison of well-being among women in a French and American city (Rennes, France and Columbus, Ohio). A major difference between French and American life, salient to anyone with a passing familiarity with the two cultures, is that the French take much more vacation than Americans (twenty-one more vacation days, on average, according to Krueger et al.). Krueger et al. are concerned with this, since, as they relate, their empirical methodology severely undersamples vacation days. However, the authors reassure the reader that

this is not a large bias. The twenty-one day difference in vacations amounts to only 5.8 percent of the year. If the U-index is 10 points lower on vacation days than nonvacation days, which is almost double the difference on weekdays and weekends, then the French U-index would be an additional 0.58 percentage points lower than the American U-index.

7. The UN Human Development Index (HDI) is a cross-national measure of well-being that, somewhat consistent with a capabilities approach, is based on normalized measures of life expectancy, literacy, education, standard of living, and GDP per capita. Among other applications, the index is used to measure the impact of economic policies on quality of life.

Although this “back of the envelope” calculation does provide reassurance that failing to monitor well-being during vacations is not a major problem, it highlights what I believe to be a more fundamental problem: national time accounting fails to properly account for time. Half of a percentage point is simply too small a marginal impact for an effect as large as a twenty-one day difference in vacation time.⁸ (Note that in contrast, the impact of moving from being a student to a nonstudent is ten times as large, and the impact of moving from the worst day of the week [Monday] to the best [Sunday] is fourteen times as large.)

National time accounting assumes that the importance of an experience is exactly proportionate to the time spent on it. However, people do not account for time in such a fashion, and for good reason. Part of the reason has to do with the attributes previously discussed, such as meaning and values, which are only crudely related to time allocation. National Time Accounting also fails to properly account for the importance of peak experiences. Episodes of strong positive and negative affect tend to be rare in most lives (see Frederickson 2000). But, while rare, such episodes tend to be significant in terms of meaning. As Kahneman’s own work on retrospective evaluation suggests, when people evaluated extended experiences, they tend to put disproportionate weight on moments of peak intensity. Kahneman views this tendency as a bias, but people themselves view it as natural. For example, people will evaluate a trip to the Grand Canyon as wonderful even if the vast majority of the time was spent on mundane, often uncomfortable, transportation—getting there, then returning home.

In fact the DRM implicitly succumbs to, and is in part rescued by, the tendency to encode experiences in terms of meaning. It divides the day into meaningful “episodes” (e.g., eating dinner, commuting to work, etc.) and then has people rate their affect during each episode. Dividing the day into such episodes reflects an implicit, if unintended, understanding that people make sense of their lives in terms of meaningful episodes and not in terms of raw numbers of minutes and hours spent in different ways. Moreover, it seems likely that what people are reporting for a particular episode is not their average affect during the episode, but some function of extremes and meaning. If, while biking to work, I get into an argument with a driver (as happens about every other week), I would evaluate the overall commute as negative, even if the altercation took place in the last few minutes of my commute and the remainder of the commute was quite pleasant. If one took

8. The problem of accounting for time would be even more serious if sleep were counted as part of the day. Currently, the denominator of the U-index does not include time spent sleeping (David Schkade, personal communication). On the one hand, this seems reasonable. If the denominator of the U-index included time spent sleeping, and one allows for an average of eight hours of sleep a day, then the .58 percentage point maximum impact of the twenty-one vacation days French-American difference would be reduced to an even more paltry .39 percentage point difference. If sleep were added to the denominator, everything other than sleep would matter even less.

time accounting seriously and did the same type of time-weighting within episodes that NTA does between episodes, it is likely that the same calculus that renders twenty-one days of additional vacation almost imperceptible would imply that such a commute was in fact a positive experience. In fact, if one really accounted for time the ways that NTA dictates, I suspect that almost everything that people care about would end up having an imperceptible impact on estimated welfare. Ironically, the validity of NTA is rescued to some extent by its failure to take its architects' own time-proportionate time accounting too seriously.

2.3.2 The Problem of Retrospection

While working on this commentary, I asked my jogging partner to report on his momentary affect. We jog together practically every day, so presumably this is an activity that we both find worthwhile. If asked to retrospectively evaluate how much we enjoy jogging, we would both rate it very highly. Indeed, when I asked him whether he was enjoying our jog *while we were jogging* his immediate response was affirmative. However, the reality of jogging is not that pleasant, and when we probed the issue more deeply he recognized that his momentary affect was really not all that positive—that he had actually been reporting his gestalt sense of the jog as meaningful, not his momentary feelings, which were quite negative. Although we usually start our jogs feeling comfortable, by the middle of the run we are almost inevitably exhausted and either too hot or too cold. And the truth is that we do not start out comfortable. I have a permanently torn hamstring that causes acute discomfort until the endorphins kick in, and my jogging partner suffers from mild asthma that is especially bad in the winter and the spring and when it is cold and when it is muggy—in short, most of the time. So why do we do it? Companionship? Health? Poor memory for pain? Probably all of these reasons and more. All I know is that I want to continue jogging with my friend, and am convinced that it is often the high point of my day, despite the misery. The Day Reconstruction Method would, in fact, reveal jogging to be a positive activity for me, but only because I would report the meaning of the activity rather than the “true” momentary affect.

Although I have no hard evidence to back the assertion, I suspect that child care is similar. Child care comes out as the second most positive activity according to the U-index, second only to socializing. Yet, again I suspect that this is because child care is meaningful and not because it is so conducive to positive emotions in the moment. Indeed, it is perhaps instructive that child care comes out *second worst* with the PATS data collection scheme. Child care comes out so inconsistently probably because the reality is not really all that wonderful most of the time but, as Krueger suggested to me in an e-mail, “no one wants to sound like they are a bad parent who doesn't enjoy being with their kids.”

2.3.3 Loss of Information with the U-Index

By encoding an activity as either negative or positive, with no finer gradations, the U-index discards a lot of potentially useful information. The stated reason for throwing out all this seemingly valuable information, according to Krueger et al., is to allow for interpersonal comparability. Summing total happiness across people is not a meaningful task, but estimating the average percent of time that people are in negative affective states is, at least in theory, meaningful. However, this implies that the U-index is effective in distinguishing between affectively negative and positive experiences, which seems questionable to me, as suggested by my anecdote about jogging.

Also, the U-index would seem to depend substantially on what specific emotions are included in the list. The U-index requires people to assess the intensity of emotions in a fashion such that intensities can be compared with one another, but it is not clear how one should compare the intensity of “happy” and “stressed.” If one is moderately happy and a bit more than moderately stressed, is that a net negative emotional state? Not to my thinking. And what if the word “happy” were replaced by “ecstatic”? As a result of this change in wording, almost certainly more events would be encoded as negative. Should the U-index depend on the implied intensity of the affective terms included in the list? Moreover, the current U-index list of six emotions includes two—“tired” and “interested”—that not only do not seem like emotions, but are not even unambiguously positive or negative.

In contrast to the coarse treatment of happiness, the U-index is very fine-grained in its treatment of time. If one takes a negative activity and makes it much more negative—for example, changing a 10 volt electric shock to a 110 volt shock, this will have no impact on the U-index, which simply encodes whether the experience is positive or negative. However, if one increases the duration of the shock by 10 percent, its contribution to the U-index will increase by 10 percent. This raises questions about the validity of decompositions presented in the chapter, such as the one that addresses the question of why older people are happier. According to this decomposition, 60 percent of the difference is due to time usage, but this conclusion seems dubious given the much greater sensitivity of the U-index to time use than to intensity of affect.

The insensitivity of the U-index to the intensity of affect is also problematic from a policy perspective. Many policies one could imagine implementing are likely to change the intensity of negative affect, but are unlikely to move people over the positive/negative line. Thus, for example, one might respond to the high U-index for adult care, not by attempting to reduce the time spent on it, but by attempting to provide assistance that would make giving such help less onerous. But, if such assistance raised negative affect

from -9 to -3 , this would have no impact on the U-index, whereas reducing the amount of time spent on adult care would have an impact exactly proportionate to the time reduction.

2.4 A Proposed Revision to NTA

The greatest strength of NTA, in my opinion, is that it evaluates well-being in terms of how people actually use their time. Although, as I have discussed, not all of well-being can be captured in such terms, how one uses time is clearly important; a life spent doing things one did not want to do is a life not worth living. The main limitation of NTA, in my opinion, is its focus on happiness, which elevates a particular hedonic feeling to an all-important role at the expense of a wide range of other things that matter, such as meaning, wisdom, and values. The specific implementation is also problematic because, as just discussed, it discards valuable information while not really achieving the interpersonal comparability that is the motivation for doing so.

There is no reason why NTA could not be improved by retaining its strengths while eliminating its weaknesses. Krueger, in an e-mail response to my verbal commentary at the meeting devoted to NTA, asked whether I thought that NTA could be improved by asking people to report whether a particular use of their time was “a waste of time.” I think it would, dramatically. Moreover, the same idea could be approached more positively by asking whether a particular use of time was a “valuable use of time.” These more general questions, I believe, come closer to measuring what makes life worthwhile than do questions that measure affect. Taking care of one’s parent may not be enjoyable, nor climbing a mountain nor jogging with a friend. But if the individuals engaging in these activities report that they are worthwhile, I believe that those individuals’ assessments of what matters to them should be accepted.⁹ Although I don’t think that such an index of whether people spend their time doing things that they want to do would be the best imaginable unitary measure of well-being, I do believe that such an index would do a better job of complementing GNP—of measuring important aspects of well-being that are not captured by GNP.

2.5 Conclusion

During the winter break of 2007, my family and I had been planning to fly to Los Angeles to go hiking in Joshua Tree National Park. However, we

9. Of course every method of elicitation has its problems, and these are almost always underestimated before one starts thinking deeply. For example, if someone makes a lot of money at a job they hate, should they respond that their work is worthwhile or a waste of time? If you are in the hospital to get needed treatment, is that worthwhile or a waste of time? Accounting is tricky, and accounting for happiness or meaning is especially so.

all felt exhausted by the prior semester and ended up staying home (and incidentally, conforming to the dictates of economic rationality by walking away from the sunk costs represented by our tickets). We had a very comfortable, relaxed holiday, catching up on sleep, friends, movies, and novels, and also work—including writing this commentary. During the entire period when we had been planning to be away, I felt happy and relieved that we had not gone to Los Angeles. I had images of changing planes, missing flight connections, looking for hotel rooms during a peak holiday season, realizing we were in the wilds without some critical piece of camping equipment, and so on. At some point, however, the irony hit me that, while I had devoted part of the vacation to writing a commentary critical of NTA, we had made exactly the choice that NTA, and particularly the U-index, would have favored—and were all very glad to have made it.

Was it the right choice? As friends returned from exotic destinations—Europe, a Caribbean cruise, Egypt—I did start to wonder. Though we had a wonderfully relaxed time, and they returned with exactly the types of horror stories the contemplation of which had helped to sustain my contentment with having stayed home, none of them regretted their decisions. In fact, I had the impression that several of them pitied us for having stayed put.

Did we make the right decision? Even if we were happier on average, I'm not sure. Moreover, I do know that if we made the same decision every time it would be a mistake, even if we spent more time in a state of happiness. Holding all else constant, it is generally better to be happy than to be unhappy. But happiness is only one of many things that make life worthwhile, and many of the other things, such as meaning, wisdom, values, and capabilities often come at the expense of happiness. Next time, I hope, we'll go hiking.

References

- Cantril, H. 1967. *The pattern of human concerns*. New Brunswick, NJ: Rutgers University Press.
- Clark, H. H., and E. V. Clark. 1977. *Psychology and language*. New York: Harcourt, Brace, Jovanovich.
- Clark, H. H., and M. F. Schober. 1992. Asking questions and influencing answers. In *Questions about questions*, ed. J. M. Tanur, 15–48. New York: Russell Sage Foundation.
- Diener, E., R. A. Emmons, R. Larsen, and S. Griffin. 1985. The satisfaction with life scale. *Journal of Personality Assessment* 49 (1): 71–75.
- Frederickson, B. 2000. Extracting meaning from past affective experiences: The importance of peaks, ends, and specific emotions. *Cognition and Emotion* 14 (4): 577–606.
- Grice, H. P. 1975. Logic and conversation. In *Syntax and semantics: Vol. 3. Speech acts*, ed. P. Cole and J. L. Morgan, 41–58. New York: Academic Press.

- Huxley, A. 1932. *Brave new world*. New York: Harper and Row. (Perennial Library Edition, 1969.)
- Karlsson, N., G. Loewenstein, and J. McCafferty. 2004. The economics of meaning. *Nordic Journal of Political Economy* 30 (1): 61–75.
- Loewenstein, G. 1999. Because it is there: The challenge of mountaineering . . . for utility theory. *Kyklos* 52 (3): 315–44.
- Loewenstein, G., and P. Ubel. 2008. Hedonic adaptation and the role of decision and experience utility in public policy. *Journal of Public Economics* 92 (8–9): 1795–1810.
- Mill, J. S. 1871. *Utilitarianism, 4th ed.* London: Longmans, Green, Reader, and Dyer.
- Nussbaum, M. 2000. *Women and human development: The capabilities approach*. Cambridge: Cambridge University Press.
- Offner, A. 2006. The challenge of affluence: Self-control and well-being in the United States and Britain since 1950. Oxford: Oxford University Press.
- Schwarz, N. 1999. Self-reports: How the questions shape the answers. *American Psychologist* 54 (2): 93–105.
- Schwarz, N., F. Strack, and H. P. Mai. 1991. Assimilation and contrast effects in part-whole question sequences: A conversational logic analysis. *Public Opinion Quarterly* 55 (1): 3–23.
- Scitovsky, T. 1976. *The joyless economy: The psychology of human satisfaction*. Oxford: Oxford University Press.
- Sen, A. 1985. *Commodities and capabilities*. Amsterdam: North-Holland.
- . 1992. *Inequality reexamined*. Cambridge, MA: Harvard University Press.
- Smith, D., G. Loewenstein, and P. Ubel. Forthcoming. Happily hopeless: Lack of adaptation to temporary versus permanent colostomy. *Health Psychology*.