QALYs VERSUS EXPERIENCE: A PERSPECTIVE FROM EXPERIMENTAL ECONOMICS

Professor Daniel Kahneman
Eugene Higgins Professor of Psychology
Woodrow Wilson School,
Princeton University, USA

14th Annual Lecture 2007

Research OHE Consulting
Office of Health Economics

www.ohe.org

ISBN 978-1-899040-68-1
August 2008
Price £10.00

QALYs VERSUS EXPERIENCE: A PERSPECTIVE FROM EXPERIMENTAL ECONOMICS

Professor Daniel Kahneman
Eugene Higgins Professor of Psychology
Woodrow Wilson School,
Princeton University, USA

OHE 14th Annual Lecture
Wednesday 13 June 2007

Office of Health Economics
12 Whitehall London SW1A 2DY UK
Tel: +44 (0)20 7747 8850
www.ohe.org

This is an edited transcript of the lecture 'QALYs VERSUS EXPERIENCE: A PERSPECTIVE FROM EXPERIMENTAL ECONOMICS' given by Professor Daniel Kahneman, Eugene Higgins Professor of Psychology, Woodrow Wilson School, Princeton University, USA on the 13 June 2007.

About the Author

Professor Daniel Kahneman

Professor Daniel Kahneman is Eugene Higgins Professor of Psychology, Princeton University, and Professor of Public Affairs, Woodrow Wilson School of Public and International Affairs. He is winner of the 2002 Nobel Prize in Economic Sciences for his pioneering work integrating insights from psychological research into economic science, especially concerning human judgment and decision-making under uncertainty. Much of this work was carried out collaboratively with Amos Tversky.

Before moving to Princeton he was a professor of psychology at the University of California, Berkeley, a fellow at the Canadian Institute for Advanced Research, a professor of psychology at the University of British Columbia, a fellow at the Center for Advanced Study in the Behavioral Sciences, and a professor at the Hebrew University in Jerusalem. Professor Kahneman is a member of the American Academy of Arts and Sciences and the National Academy of Sciences. He is a fellow of the American Psychological Association, the American Psychological Society, the Society of Experimental Psychologists, and the Econometric Society. He has been the recipient of numerous awards in addition to his Nobel Prize, among them the Distinguished Scientific Contribution Award of the American Psychological Association, the Warren Medal of the Society of Experimental Psychologists, and the Hilgard Award for Career Contributions to General Psychology.

Professor Kahneman was born in Tel Aviv but spent his childhood years in Paris, France, before returning to Palestine in 1946. He received his bachelors degree in psychology (with a minor in mathematics) from Hebrew University in Jerusalem, and in 1954 he was drafted into the Israeli Defense Forces, serving principally in its psychology branch. In 1958 he came to the United States and earned his Ph.D. in Psychology from the University of California, Berkeley, in 1961.

During the past several years, the primary focus of Professor Kahneman's research has been the study of various aspects of experienced utility (that is, the utility of outcomes as people actually live them).

Office of Health Economics

The Office of Health Economics (OHE) was founded in 1962. Its terms of reference are to:

- commission and undertake research on the economics of health and health care;
- collect and analyse health and health care data from the UK and other countries;
- disseminate the results of this work and stimulate discussion of them and their policy implications

The OHE is supported by an annual grant from the Association of the British Pharmaceutical Industry and by revenue from the sales of its publications, consultancy and commissioned research.

Introduction

I know essentially nothing about quality adjusted life years (QALYs). The little I know I learned from Professor Paul Dolan, Chair in Economics at Imperial College London, when he was a visiting fellow at Princeton University in 2005. I have more experience of a closely related topic, known as "contingent valuation", which raises a very similar problem. Contingent valuation is a survey technique used to measure the value of public goods, where people answer questions such as: "How much would you be willing to pay to preserve a certain species, or to achieve certain environmental goals?" The contingent valuation technique was introduced by economists as a tool for economic analysis of costs and benefits. It is quite extensively used both in the US and all over the world as a measure of the value of public goods.

I happen to be a rather severe critic of contingent valuation. The reason I am a critic is because I think I know something about survey methods. What you get when you elicit willingness to pay is not a measure of utility, because the responses fail to satisfy the axioms and the logical requirements of a proper measure of utility. I am afraid that something very similar may be happening with respect to QALYs.

Concepts of utility: decision utility and experience utility

Let me begin by introducing a distinction between two classical concepts of utility. The first one, as it has been used in economics over the last 100 years or so, is "decision utility". It is a quantity, a value that is inferred from choices and is, in turn, used to explain choices. It is all about what people want and what people choose. That is what the axioms of expected utility theory are about: people's preferences for outcomes, whether certain or uncertain. I believe this is the standard meaning of utility within modern economics.

The other meaning of utility, however, goes back to Jeremy Bentham[1] who introduced the concept. As Jeremy Bentham used the term, utility was a measure of experience; it was a measure of pleasure and pain. This sense of utility, which was completely different to the first mentioned above, has survived – but only in philosophy and not in economics. Philosophers think of utility as an aspect or a measure of experience, quite distinct from decision utility, which is a measure of desire or wanting. This is an important distinction.

QALYs are inferred from preferences and are therefore measures of decision utility. What the corresponding experience utilities are is an open question. My view is that there are many reasons to doubt the value of QALYs as predictors of the actual experience of health states.

Another much debated question is "Whose decision utility?" i.e. whose preferences are we trying to satisfy? In terms of healthcare, are we interested in the patients' preferences? i.e. are we interested in what they believe about their state of health and what trade-offs they would be willing to accept between remaining in their state for a longer, or a shorter time in

[1] J. Bentham, *An Introduction to the Principles of Morals and Legislation*, (originally published in 1789) ed. J.H Burns and H.L.A. Hart (The Athlone Press, 1970), pp. xliii, 343. Reprinted in paperback with new introduction by F. Rosen (Clarendon Press, Oxford, 1996), pp.cxii, 343.

perfect health, and so on? Or is it the public's preferences? In the UK, as in most other places, it is the public that is consulted about states of health. There is no guarantee that public preferences are the same as patients' preferences, and indeed we *know* that there are substantial discrepancies between the QALYs that are inferred from surveys of patients or of the healthy public.

Decision utility: patient versus public preferences

We should know – and here I am interjecting some psychology – what it is that those QALYs measure. Suppose a member of the public is asked about a preference between a certain state of health, say being blind, and perfect health, and is offered a standard gamble or a time trade-off. I would argue, as a psychologist now, that in this situation we are measuring fear. The people who engage in that exercise are not blind right now, they are not experiencing anything. The thought experiment they are being asked to do is, in effect, the following: "Consider being blind. How afraid are you?" What we get is translated from fear and, ultimately, whatever it is that the respondents are expressing is, I believe, a measure of fear.

Unfortunately, there is no guarantee that the feat that healthy people experience corresponds to the utility they would experience if they were afflicted. Indeed, there are systematic discrepancies between anticipatory emotions such as fear and the feelings or emotions that are actually experienced. In particular, it is well established that the public is generally more afraid of disease than the patients who are actually experiencing it. In other words, patients who experience states of less than perfect health attribute higher utility to their state than the public would. This raises a very significant question: who are we to take seriously when there is a systematic conflict between the public and the people whose life is assessed?

There are more specific discrepancies between the utilities of the public and the utilities of patients. When people report on their own experience, feelings of energy or vitality are extremely important to their self-assessed well-being. In contrast, the dimension of vitality has distinctly less weight in the public valuations of states of health. This is one example of several systematic discrepancies between public utilities and patient experience.

So far, I have drawn several contrasts between, first, experience utility and decision utility and second, between patient utility and public utility within decision utility.

Prospect Theory

I now want to draw attention to another problem. The logic of utility theory can be crudely summarized as follows: if there is a preference order and if the preference order conforms to certain axioms, then it is possible to assign utilities to outcomes. We demonstrate that utility is measurable by testing whether the axioms are satisfied.

However, what if the conditions are not fulfilled? What if there are no stable preferences? What if there is no stable preference order? What if the preferences that we observe systematically violate the axioms of utility theory? In that case, strictly speaking, there is really no such thing as utility. I believe that we are in that situation; preference orders are

unstable and where the axioms are predictably and systematically violated. The existence of what we are trying to measure is therefore in doubt.

I can use a relevant parallel situation to illustrate one problem we have here – something of an extreme example but nevertheless relevant. Physicists used to believe in the existence of ether. They no longer do. However, if you are intent on measuring some property of the ether, how do you deal with the possibility, or the fact, that ether does not exist? I suspect we are in a similar situation when we are trying to measure QALYs.

I speak from the perspective of somebody who has studied decision-making and the way that individuals make decisions. The theory that I am most closely associated with, and which I wrote with my late colleague Amos Tversky – who died before the award of the Nobel Prize, which he would certainly have shared – is called "prospect theory". The core idea of prospect theory, which had been anticipated by Harry Markowitz, is that the carriers of utility are not what they are assumed to be in standard utility theory.

In standard utility theory, the carriers of utility are final states, the states that will exist after the uncertainty is resolved. In prospect theory, in contrast, the carriers of utility are changes relative to a reference point. Let me illustrate what this means. Suppose we that someone has a choice between owning, say, £2 million for sure, or equal probabilities of owning £1 million or £3 million. Which of these two prospects has higher utility? In expected utility theory, this question can be answered without asking for more information. In fact, it is very important to know another fact: what is the chooser's wealth now? We know, for example, that a decision maker is much more likely to prefer the gamble if her current wealth is 3 million than if she currently holds 1 million. We need an extra parameter to predict preferences. In standard utility theory, where the carriers of utility are final states, the reference point does not matter. The prospects look the same and have the same value regardless of the perspective from which you look at them.

Prospect theory includes the missing parameter, the reference point relative to which gains and losses are assessed. This can make a very substantial difference. One of the robust findings in the study of preferences is that people tend to be risk-seeking when they face a choice between a sure loss and one that is merely probable. In the example above, the outcome of owning 2 million for sure is evaluated as a loss when current wealth is 3 million. The convexity of the value function in the domain of losses implies a preference for the gamble. If the reference point is 1 million, of course, most people will prefer the sure thing over the gamble.

Similar results are found in the domain of health states: the reference point matters. Indeed, Ubel, Loewenstein and their colleagues recently make that point in an impressive set of experiments[2]. They studied both the experience utility and the decision utility of colostomy in several groups, including a group of patients whose colostomy had been reversed. They therefore had normal people who judged what a colostomy would be like and also reported on their current (normal) their own experience; they had colostomy

[2] D.M. Smith, R.L. Sherriff, L. Damschroder, G. Loewenstein and P.A. Ubel (2006) "Misremembering colostomies? Former patients give lower utility ratings than do current patients" *Health Psychol*. 2006 Nov ;25 (6):688-95.

patients; and they had post-colostomy patients, whose colostomy had been reversed. The study yielded several striking findings. First, there is little if any difference in experience among the three groups. I find this almost impossible to believe, and my difficulty in believing this true fact illustrates the main point of the experiment. People expect to find an enormous difference in reported mood between colostomy patients and non-patients, but in fact there is little or no difference. Most of us would intuitively predict a colostomy is devastating, but the mood results contradict that expectation. The second main finding is that people who have a colostomy remember their past as much happier than it actually was (inferred from the healthy controls). Thus, they remember themselves as being very happy before the colostomy but their mood is currently normal.

The critical result, which strongly supports the importance of the reference point, refers to how people post-colostomy evaluate their state while they had the colostomy. The answer is straightforward: they retrospectively think they lived through hell. So we have a very strange combination of results: people living in the state of having a colostomy not only cope with it but they cope with it reasonably cheerfully. Prior to that, they are very afraid of it; after that, they hate the thought of it.

The experimenters also measured QALYs, using time tradeoff, and they found more support for the reference point: the reduction of QALY associated with a colostomy was twice as large among the post-colostomy patients than among the current colostomy patients.

These are the kinds of discrepancies that we have to live with when we think of what is the utility of a colostomy. How do we put it all together?

I do not suggest that we stop using the term "utility". The term is far too useful to be given up easily. We will certainly continue to talk of decision utility and experience utility. However, we must remember that our problem in dealing with the utility of health states is conceptual – it is not a measurement problem. What we are trying to measure is not a single thing, it is a collection of related things. It should be obvious that the differences between different ways of looking, say, at the utility of a colostomy are too systematic to be ignored but there is no easy way out of the difficulty. What we know with certainty is that when the problem is eventually sorted out, the solution will be much more complex than the standard story.

When I introduced a label for the distinction between decision utility and experience utility, a decade ago, I was intrigued by the possibility of reviving Bentham's concept of experience utility[3]. The focus of my efforts was time. I found out something that students of QALYs have known for a while: the only way to measure experienced utility is in units of time.

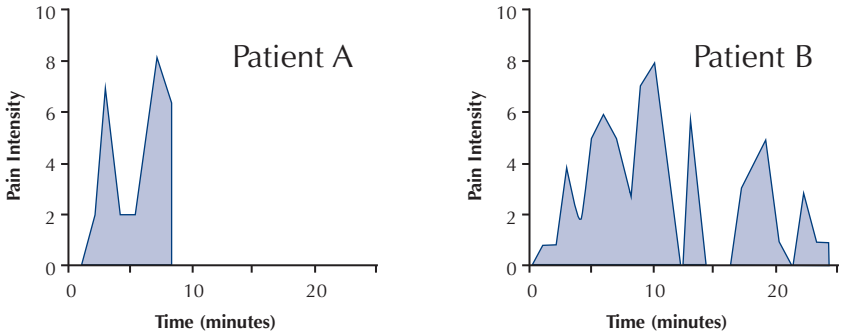
With QALYs we measure utility in units of time. It turns out that, in general, this will be the way in which we *have* to measure experience utility too. This ultimately boils down to a simple point. Starting from a given negative experience, we have at least these two ways of making it worse: we can make the discomfort more intense or we can make it longer. There

[3] D. Kahneman, P.P. Wakker and R. Sarin (1997) "Back to Bentham? Explorations of experienced utility." *Quarterly Journal of Economics*, 112, May 1997, 375-405.

must be some equivalence between these two manipulation. We know how to measure time, so we end up calibrating pain on a ratio scale in units of clock time. Note this is identical to QALYs: five minutes of pain at Level 2 are as bas as ten minutes of pain at Level 1 because the integral is the same in both cases.

I now want to show you some results from an example where patients are undergoing an unpleasant experience: a colonoscopy (Redelmeier, Katz and Kahneman, 2003)[4]. Figure 1 shows the profiles of pain of two patients (both males) who were asked every 60 seconds to report their pain on a zero to ten scale, where ten is intolerable pain and zero is no pain. These are two different patients, but you could think of it – because for my point it will be the same – as two colonoscopies of the same patient at different times.

Figure 1: Two patients' profiles of pain experiencing a colonoscopy



Source: Redelmeier and Kahneman, 1996

We want to assess how much the two patients suffered and who of them suffered more, from the perspective of an objective observer seeing this profile. Every observer will agree that Patient B had a worse colonoscopy than Patient A: B's colonoscopy lasted 22 minutes while A suffered for only eight minutes. Furthermore, if you consider the cumulative distribution, you will find that B had more pain than A at every level. This is dominance: Patient B had a worse time than Patient A.

What we are doing here has a long history: in 1881 Edgeworth spoke of a hedonometer (Edgeworth, 1881)[5] which would record the quality of the experience at every moment. He argued that happiness over a period of time is the area under the curve. This is true, of course, only if momentary utility is measured on a ratio scale. We cannot assume that the ratings of pain that the patients offered are on such a scale, where a rating of 4 is truly twice as bad as a rating of 2. Together with some colleagues, I discussed the conditions under which the original ratings can be subjected to a monotonic transformation into a ratio-scale measure of utility (Kahneman, Wakker and Sarin, 1997). After that transformation, the temporal integral of moment utility is an acceptable measure of the

[4] D.A. Redelmeier, J. Katz and D. Kahneman (2003) "Memories of colonoscopy: a randomized trial." *Pain*, July 2003, 104(1-2), 187-194.

[5] F.Y. Edgeworth (1881) "Mathematical Psychics: An essay on the application of mathematics to the moral sciences". London: C. Kegan Paul and Co., 1881. Available online at http://socserv2.mcmaster.ca/~econ/ugcm/3ll3/edgeworth/mathpsychics.pdf

total experienced utility over a period of time – but the conditions are very stringent.

Remembered utility

Let me now introduce another way to measure the utility of an episode. Suppose that we ask the colonoscopy patients, when the procedure is done, the following question: "How bad was it?" I will call their rating of their own experience a measure of "remembered utility". We are not measuring the quality of experience, because the experience is over by the time we elicit an evaluation. What we measure is how much these patients *think* they suffered.

Surprisingly, Patient A evaluation of his experience was substantially worse than B's evaluation of his. The reason was simple: Patient A had been unlucky, because his colonoscopy ended as he was was experiencing very severe pain. In contrast, Patient B's colonoscopy terminated gently, with pain at a low level. We learned from this experiment and others that we can predict ratings of remembered utility with good accuracy from the characteristics of the profile of moment utility. A simple rule that does a good job is the "peak and end rule". An unweighted average of the worst pain (the peak) and the pain at the end of the episode is a good predictor of the individual's subsequent evaluation of the episode as a whole.

Note that one parameter of the experience is obliterated in this equation: its duration. The phenomenon, which has been observed repeatedly, is called duration neglect[6]. When people evaluate episodes of pain – we have a few cases in which they evaluate episodes of pleasure as well – the duration of the episode does not seem to matter. When people look back on an episode, they perform a simple operation on the remembered experience of a few selected moments, which yields a representative moment – and they evaluate the entire episode from the utility of that representation. People will also make choices according to their remembered utility. Indeed, it is possible to create a situation where people will expose themselves willingly to more pain rather than less pain, because they follow the "peak and end rule"[7].

Let me illustrate this by an experiment that we actually conducted. We applied a procedure that is commonly in psychology laboratories to inflict a moderate amount of pain. Participants in this cold-pressor procedure stick their hand, up to the wrist, in unpleasantly cold water. We were quite humane, and the temperature of the water was 14^0C. This temperature is tolerable, but it gradually becomes painfully cold. At the end of one minute of it, people report pain levels of, say, seven to eight on a scale of ten.

We had the same people undergo two cold-pressor trials. The 'short' experience consists of 60 seconds at 14^0C, at the end of which the subject is invited to remove his hand from the tub and is given a warm towel. The 'long' experience begins with the same 60 seconds, at which time the subject is told nothing at all; the experimenter silently opens a valve that lets in a small amount of warmer water so that, over the next 30 seconds, the temperature of the water in the tub rises from 14^0C to 15^0C.

[6] B.L. Fredrickson and D. Kahneman (1993) "Duration neglect in retrospective evaluations of affective episodes." *Journal of Personality and Social Psychology*, 65, 45-55.

[7] D. Kahneman, D.L. Fredrickson, C.A. Schreiber, and D.A. Redelemeier (1993) "When more pain is preferred to less: adding a better end." *Psychological Science*, 4, 401-405.

Participants had those two experiences in randomised order, separated by seven minutes. Seven minutes after the second trial they were brought in and told, "You had two experiences: one with your left hand; one with your right hand. For your third trial (they had been told to expect three) we are going to repeat exactly one of the two experiences that you just had. Which would you rather repeat: your left-hand experience or your right-hand experience?" Depending on the details of the situation, up to 80% of people chose the long experience rather than the short experience. These preferences are compatible with the Peak/End rule: the average of pain at the peak and at the end of experience is actually milder for the long experience. The short trial ends at the moment of maximal pain – much as it did Patient A in the colonoscopy study. The long trial resembles the experience of Patient B.

The conclusion is straightforward, and it appears to be quite general: when people make a choice of whether to repeat an experience, they will evaluate it by its remembered utility, which in turn is determined according to the Peak/End rule.

We face a deep conceptual problem construing experience utility, just as we did for decision utility. The problem is different, but not easier. We find measuring the total experienced utility of a bad episode as the integral of momentary pain over time makes a great deal of sense. However, it also seems reasonable to consider how the subjects themselves remember their experiences. Indeed, remembered utility is especially important because it determines people's preferences for the future. The difficulty, of course, is that remembered utility is not a measure of what actually went on; it is systematically biased toward a neglect of duration and an overweighting of Peak and End. People's choices are susceptible to exactly the same bias.

We all wish for a single number that could serve as a sufficient statistic to describe the utility of a subjective experience, but no such number exists.

Measurement of well-being – the "Day Reconstruction Method"

I committed myself a long time ago to the idea of total utility; that is, to measuring the utility of episodes by the temporal integral of momentary experienced utility. It seemed to me that the people – such as Patients A and B – who apply the peak-end rule are making a mistake. I believed that if we know the temporal profile of an experience has been, we are in a better position to evaluate it than the subject himself.

This idea has implications for the measurement of well-being. At least in principle, we can measure well-being in two quite different ways. Research on well-being has traditionally relied on surveys in which people report their satisfaction with life. Clearly, this is equivalent to a measure of remembered utility. I call it the happiness of the remembering self. But it is also possible to measure "experienced happiness" by collecting reports of immediate experience, which involve neither memory nor much evaluation. This is what is done in the so-called "experience sampling technique".

My colleagues and I developed an alternative method for evaluating experienced well-being,

which we call the "day reconstruction method" (DRM). Participants in such a study are not asked for a global evaluation of their life. Instead, they provide a detailed description of one day, which they break into episodes, or scenes. We ask them to report for each episode what they did, who was with them, and how they felt. The DRM provides detailed information about time-use. In addition, it allows us to compute a duration-weighted integral of experienced utility, just as Edgeworth would have us do it[8].

We have developed a summary measure that I believe to be promising. We call it the "U-index" – for "unpleasantness". We examine episodes one at a time. Our criterion for calling an episode "unpleasant" is that the most intense negative feeling that is reported for that episode is more intense than the maximum of the positive emotions. In our original study the negative emotions included anger, depression, and tension. The positive emotions included happiness enjoyment and friendliness. The choice of emotions will vary in different applications. The data suggest that, so long as there is a sufficiently large number of rating scales, overall estimates will be roughly the same. We proposed the U-index as a possible measure of experienced well-being – and as one element of a more comprehensive assessment of well-being.

Next, some miscellaneous facts about the U-index. First, among American women, the average is about 18%. This means that 18% of the time on average – and we have large samples for which this was measured – American woman report being in a negative state. The U-index is lower in Denmark, where it is about 14%. The average day for the Danes is as good as the weekend for the American women.

Second, there is an extremely uneven distribution of the U-index. Pain and suffering are concentrated in the population. We have calculated how much of the total time spent suffering (in the population) is contributed by the 10% who suffer the most. The estimate is close to 40%. In other words, the top 10% of sufferers account for 40% of total suffering. This is almost certainly an underestimate of the true inequality in the distribution of pain, because the people who suffer a great deal of the time also suffer more intensely.

The distribution of the U-index has immediate policy implications, if one of the objectives of policy is to reduce the incidence of suffering in society. To achieve this objective you should probably focus first on mental illness, and on conditions such as depression and chronic pain, which cause relentless suffering.

As a secondary implication, you might want to improve people's lives by changing the way they spend their time, but these effects will be small – reflecting the general fact that the variance of well-being that is due to life circumstances is small (at least within prosperous societies) relative to the variance that is due to individual differences in character and personality. If you completely removed commuting from people's lives – and commuting is one of the worst parts of a day – you would reduce the U-index by maybe 1½ per cent. However, a reduction of the U-index from 18% to 16.5% would actually be a major achievement. You should think of it as reducing the amount of suffering in society by 10%.

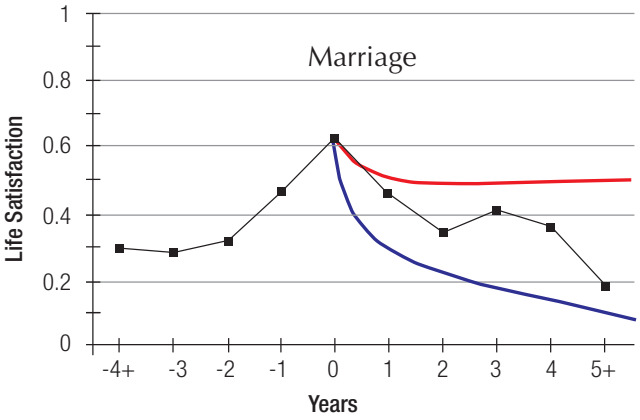
[8] D. Kahneman, A.B. Krueger, D. Schkade, N. Schwarz and A. Stone (2004) "A survey method for characterizing daily life experience: the day reconstruction method." *Science*, 306, 1776-1780.

It is a goal that we should certainly strive for.

Adaptation

I now present the emblematic results of well-being research, which are taken from the German Socio-Economic Panel, a longitudinal survey of households in Germany [http://www.diw.de/english/soep/26636.html]. The data I present here concern life satisfaction among women before and after marriage. People are asked: "How satisfied are you with your life?" Figure 2 shows some interesting results.

Figure 2: Competing interpretations of adaptation: marriage



Source: Clark, Diener and McCulloch (2001)

You could explain this result in two quite different ways. One is that people get used to the bliss of getting married. They may still be in bliss but now they expect to be in that state, so the level of happiness at which they declare themselves satisfied with their lives changes. This is illustrated by the red curve in Figure 2, where affect, or the hedonic experience, adapts less than life satisfaction. The second hypothesis, as illustrated by the blue curve in Figure 2, is that affective experience adapts more quickly than life satisfaction.

The research was conducted by an interdisciplinary team, which included the economist Alan Krueger -- his name is probably well known to many in the audience, and he has now taken over this project. We started out convinced that the red curve would describe the data. But all the indications we have now – I would say that the evidence is almost conclusive – are that the blue curve describes the data. We now believe that adaptation to a new circumstance is likely to occur sooner in affect or emotion than in life satisfaction.

The natural way to think about this issue is to ask: "How could it be? Why does marital bliss subside?" This sends us off searching for an explanation of the drop. We take it that the drop is the phenomenon that requires an explanation. We want to know why people adapt to the state of being married so that they no longer enjoy it?

However, there is a different way of looking at the same facts. We have analysed how women who are married and women who are not married spend their time. The day reconstruction method that we apply is, among other things, a study of time use. We learn from it what people do with their time, how much of that time they spend on different activities. Table 1 shows the difference in time-use between women who do and do not have a mate.

Table 1: How women with a mate and with no mate spend their time (US)

% time	No mate	Has mate
Alone	28%	15%
Immediate, nuclear family	17%	38%
Intimates (friends, family)	16%	3%
Compulsory home tasks	23%	31%
Discretionary activities	41%	35%

Women who have a mate spend much less time alone – and people do not like being alone. Score this as a clear advantage for being married. They spend much more time with their immediate nuclear family, but it is not at all clear that this is a benefit, because married women spend much less time with friends and intimates, and that is a big loss. They also spend more time doing chores that they do not enjoy doing, and less time in discretionary activities that they enjoy rather more.

When you look at the balance in terms of how people spend their time, I submit that the mystery of why people are not much happier after than before marriage effectively disappears. The reason is that married life involves a mix of costs and benefits. What really requires an explanation is perhaps not why the "before" and "after" states are so similar. The real mystery is why people are so happy getting married if most of them will soon be in much the same state as before. We have an explanation for that as well.

Many good things happen during the months and years that surround the wedding. There is a new intimacy and people think a great deal about that aspect of their life. During that period, relations with one's mate are a constant preoccupation, which colours one's view of life. Over time, of course, people will start thinking about other things.

I propose that our enjoyment of any moment of our life is normally determined by the immediate local context, where we are, who we interact with, what we do. The immediate context dominates our experience because the immediate context is mostly what we think about. However, there are exceptional times of life – especially times in which life changes

a great deal – during which our thoughts are not dominated by the local context. If you are very much in love, you can be happy while stuck in traffic! If you are bereaved and suffering, you can be miserable at a feast. This idea explains is the bump of life satisfaction in the figure, rather than the drop from the peak back to the normal state.

I strongly recommend considering time-use for anyone who wants to understand well being. Note that the focus on time-use draws our attention to the steady-states of living with or without a mate, rather than on the transition from one state to the other. If you focus on the steady-state, the mystery of adaptation is no longer a mystery.

More generally, I wish to suggest that if you want to understand well-being, including the effects of health states on well-being, you will want to know what people are attending to, because what people think about is what controls their emotional state.

Adaptation is to be understood in large part as a change in the allocation of attention. When someone has recently had an accident that made him (or her) paraplegic, you know that during the early period the individual will hardly think of anything else but their tragedy. Within a month, however, paraplegics are in a reasonably good mood more than half of the time. The point is that being a paraplegic (as in most other health states) is best seen as a part-time situation. You are a paraplegic when you are thinking about it. You are a paraplegic when you are engaged in some activity that, because of your state, causes you pain or discomfort. When you laugh at a joke, when you enjoy a meal, when you talk with friends, when you get irritated by the Prime Minister, being a paraplegic does not necessarily make you different from other people.

The focusing illusion

Some time ago, my colleague David Schkade and I were challenged to study the question of whether people in California were happier than they were elsewhere[9]. Our casual observation was that most Americans believe that people are happier in California than in other parts of the country, but we suspected this belief could be wrong. We investigated two separate questions: "Are people happier in California?" and "Do people think that people are happier in California?" The results were quite clear: people are not happier in California – at least by our measures – but most people erroneously believe that residents of California are happier. The people who live there and the people who live elsewhere share the same belief.

Why does this occur? It is obvious to every American that California has a much nicer climate than most of the rest of the country. And climate is the first thing most people think about when California comes to mind. However, the effect of climate on life satisfaction turns out to be negligible. Evidently, people do not spend much time thinking about their climate – or thinking about the fact that they live in California rather than elsewhere. When a question reminds you to think of California, your attention is drawn to its superior climate, and by focusing on this aspect of life you end up exaggerating its

[9] D.A. Schkade and D. Kahneman, "Does living in California make people happy? A focusing illusion in judgments of life satisfaction". Psychological Science, 1998. 9: p.340-346.

importance. This is the focusing illusion, which David Schkade and I summarized by a maxim: "Nothing in life is quite as important as you think it is while you are thinking about it"[10].

The focusing illusion creates serious problems for the measurement of QALYs, because the procedure of measuring the anticipated utility of a state of health inevitably causes the respondents to focus on that state of health. They will then exaggerate its importance relative to other aspects of life.

Norbert Schwarz, another member of our team, asked people the question "How much pleasure do you get from your car?" He also obtained details about the model of the car the respondents drive. Americans refer to the well-known Blue Book, updated every year, which provides an estimate of the dollar value of used cars, defined by model, year and condition. Schwarz and his students were therefore able to compute the correlation between the pleasure they report having from their car and the monetary value of the car. They found a correlation of respectable size (0.35)[11].

The research design also included another question, which can be paraphrased as follows: "Remember your commute this morning. How much did you enjoy it?" As you probably guessed, the correlation of the actual pleasure of commuting with the value of the car is zero.

When do we get pleasure from a car? The answer is that we get pleasure from a car when we think about it. Some of us think of our car rather frequently, and these people will get more pleasure or more pain from it. Most of us think about our car (and draw pleasure that is correlated with its dollar value) when riding the car for pleasure – but these occasions are rare.

Another point adds to the power of the focusing illusion – and to the biases it causes. Think of someone who lives in California, and try to assess how happy that person is. You probably were reminded of the wonderful climate of California. But you did more. Without intending to do so, you almost certainly imagined someone living in California who was thinking about living in California. Because Californians rarely have these thoughts, your estimate of their satisfaction with life is bound to be biased. Californians do not think much about where they live. They just live there.

The focusing illusion is significant to QALYs because it is built into the very operation of measuring QALYs. The QALY question will focus respondents' attention on what it is like, for example, to live with a colostomy. The resulting focusing illusion will influence patients, by drawing attention to an aspect of their life that they often mercifully forget. Members of the public who estimate the utility of the same state will imagine the state of a patient who has a colostomy and who is currently thinking about it. People who have a colostomy certainly think about it often, and not happily. But they mostly think of other things. We

[10] D. Kahneman, A.B. Krueger, D. Schkade, N. Schwarz and A.A. Stone (2006) "Would you be happier if you were richer? A focusing illusion." CEPS Working Paper No. 125, May 2006, Princeton University Centre for Economic Policy Studies.

[11] Xu, Jing and Norbert Schwarz (2007), "How do you feeling while driving your car? Depends on how you think you about it," unpublished manuscript, University of Michigan.

can infer from the fact that their average mood is not very low is that they do not constantly think of their colostomy. The same is true of paraplegics and of many other conditions.

There are exceptions to this analysis of adaptation and focusing. These are conditions to which people do not adapt because they cannot think of other things. We know that this is the case for both chronic pain and depression. High ambient noise is another situation to which people do not adapt very well. What is striking about these three exceptions is that they all implicate attention. The primary function of pain is as an attention signal. The essence of depression is that sufferers are trapped in a cycle of negative thoughts that they cannot break. Like pain, noise is a potent attention signal. In all three conditions the normal process of withdrawing attention from a steady situation is prevented.

Some years ago I believed in duration-weighted experience utility as the measure of well-being. My life was relatively simple when I believed that. However, I have found that this position is not tenable and have been forced into what I now call a hybrid model of well-being. In the hybrid model both experienced happiness and life satisfaction matter. In other words both experienced utility and decision utility matter. Experienced happiness is duration-weighted, and is measured, for example, by the U-index. Life satisfaction or some other measure is a measure of what people think about their life when they think about it. Experienced happiness and reported life-satisfaction are very imperfectly correlated – the correlation coefficient is 0.4[12]. If you correct for measurement noise, you might get up to 0.6. The two are irreducibly different, and I do not think either of them is expendable. We must measure them both. This conclusion complicates the idea of measuring the utility of health states, because in a hybrid model both living in the state and thinking about it matter.

One of my goals in this lecture was to draw up a list of specifications for an adequate measure of the utility of health states. Unfortunately, the list is not very helpful. I found that there are reasonable grounds for measuring both decision utility and experienced utility. I found reasonable grounds for measures that are duration-weighted and others that are not, for measuring patients' values and for eliciting values from the public. Life is complicated. Perhaps the first step we need to make is to admit that what we have been trying to measure does not exist. We have been looking for a single measure of utility, but this search is bound to fail. There are multiple concepts of utility, each with some some normative status, there must be trade offs between them, and there is no obvious basis for setting these tradeoffs. Some people who admit to the complexity wish to ask people a single question and have them deal with the trade-offs by themselves. I find this approach unconvincing, and for the moment would favour eliciting multiple measures – and thinking hard about what to do with them all.

[12] D. Kahneman and A.B.Krueger (2006) "Developments in the measurement of subjective well-being." *Journal of Economic Perspectives*, 20 , 3-24.

Conclusions

In conclusion, the following is a list of the things that we would like a measure of the utility of health states to satisfy:

- It should be the same for the public and for patients, because if we get a measure where there is a substantial discrepancy between what patients feel and what the public wants, we have a problem;

- It should correspond to the actual utility of health states as experienced by patients;

- It should correspond not only to what the patients experience but also to what they want. This distinction is necessary because sometimes people want things that they will not enjoy and do not want things that they will enjoy. The correspondence between wanting and enjoying, between wanting and experiencing, which was taken for granted in many analyses, is actually very far from perfect;

- It should conform to the axioms of utility theory;

- It should be easily and sensibly translated into monetary units.

I have argued that what we are trying to measure here does not exist. There is no such thing that satisfies all these requirements, and it would be better to admit that the problem is conceptual than to go on pretending that all we face are issues of measurement error.

Ultimately, the issue of how to measure health states is a policy issue, which has to be decided as other policy issues are decided. The decision utilities and the experience utilities we measure will be are inputs to that decision. Once we admit that there is no single number that is the utility of a health state, someone has to make the difficult choices. I would lay that problem at the feet of the policy decision-maker rather than pretend that survey respondents can solve it.

Details of all OHE publications, and how to order them, can be found on the OHE website (www.ohe.org) or by contacting:

Office of Health Economics
12 Whitehall
London
SW1A 2DY
UK

Tel: +44 (0)20 7747 8850
Fax: +44 (0)20 7747 8851