Distinguishing between Experienced Utility and Remembered Utility

Adam Oliver*, Department of Social Policy, London School of Economics and Political Science

*Corresponding author: Adam Oliver, Department of Social Policy, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK. Tel.: +44-(0)20-7955-6471; Email: a.j.oliver@lse.ac.uk

In his 2015 book, *Valuing Health*, the philosopher, Daniel Hausman, in referring to experienced utility maximization, touches on the question of whether people accept, and ought to accept, the assumption of health maximization vis-à-vis their own lives. This essay introduces Hausman's arguments on experienced utility, before outlining the intellectual catalyst for the renewed interest in the maximization of experienced utility as an appropriate ethical rule; namely, the literature that arose in the 1990s that demonstrated that due to the so-called gestalt characteristics (e.g. the heavy emphasis that people place on how moments in an experience are ordered), an individual's retrospective and prospective assessments of an event—i.e. the remembered and decision utilities associated with that event—often systematically differ from the utility that he/she experiences or can reasonably expect to experience during the same event. The essay then touches upon some of the implications that this debate has on assuming that the maximization of quality-adjusted life years is the best way to judge the relative worth of different health profiles. To conclude, it is argued that, although contextual, it is sometimes important to consider in the policy discourse the apparent effects of the gestalt characteristics on remembered and decision utility.

Introduction

Having taught a postgraduate course titled *Valuing Health* for more than a decade, there is much with which this author can sympathize in Daniel Hausman's (2015) book of the same name. For instance, the confidence that many within the health policy community have in the ability of the commonly used health state value elicitation instruments, such as the rating scale, the time trade-off, the standard gamble and the person trade-off (which are all used to generate the quality weights used in quality-adjusted life years, or QALYs—for a review, see, for example, Drummond *et al.*, 2005), to generate interpersonal cardinal health state values is misplaced. However, this essay is not, for the most part, focused directly upon the valuation of individual health states, but rather is motivated by extending the discussion on another conundrum to which Hausman alludes, i.e. whether the value or well-being associated with the utility of a temporally extended profile ought to equate to the sum of the moment-by-moment experiences of utility of which that profile is comprised.

The essay begins by outlining Hausman's critique of those who wish to identify well-being with the experienced utility of events. The essay will then consider some of the evidence and arguments that gained currency in the 1990s, that demonstrated that an individual's relative retrospective/prospective assessments of different events (which indicate their remembered/decision utility of those events) often differ systematically from the relative utility that he/she has experienced, or can expect to experience, from the same events. Some attention will be paid to the possible implications of the differences between experienced utility and remembered/decision utility for the broad acceptance of QALY maximization as the appropriate normative rule in healthcare decision-making, an issue that has thus far received little consideration in the health policy discourse. The essay ends with some concluding thoughts.

Hausman's Critique of Experienced Utility Maximization

Modern welfare economics is based upon the assumption that people will choose—and ought to choose—among competing alternatives so as to maximize the utility that they expect to experience. That is, it is

doi:10.1093/phe/phw014
Advance Access publication on 28 April 2016

assumed that decision utility should and will equate to experienced utility. Many scholars maintain, however, that expectations of experienced utility, and thus preferences, are unreliable (Kahneman *et al.*, 1997). Hausman, in referring to the work of the Nobel Laureate, Daniel Kahneman, and his many collaborators, and, in particular, Dolan and Kahneman (2008), reflects on the proposition that well-being is not predicted by preferences (and thus is not guided by decision utility) and instead ought to be equated with the subjective value of the instant or momentary states that comprise an event or episode, as those events are lived. The integral of all instant moments of utility gives the total experienced utility of an event, or, in other words, the net sum of hedonic happiness and suffering, an approach that mirrors Benthamite utilitarianism. The advocates for this approach argue that we ought to seek to direct our activities such that experienced utility is maximized, a conjecture that Hausman notes has drawn serious interest from several governments.

In extending the argument specifically into the domain of health, Hausman maintains that Dolan and Kahneman favour experienced utility evaluation in health policy because they assume that (i) well-being is hedonic happiness, and (ii) the value of health consists in its contribution to well-being (and hence happiness). Thus, Dolan and Kahneman express the view that it is better to assess how health conditions affect actual experiences of life than to get individuals to predict the impact of these conditions, and propose that elicitation methods that may best uncover instant experienced utility, such as the so-called day reconstruction method or participant use of electronic diaries to periodically report current mood, are the most appropriate to use. To summarize, Hausman (2015: 107) notes that 'Dolan and Kahneman conclude that to, "represent the effect of different health states on people's well-being more accurately, we propose that economists in health and elsewhere shift their attention from the measurement of decision utility towards the measurement of experienced utility."'

Vis-à-vis valuing health, Hausman is in agreement with Dolan and Kahneman's antagonism towards relying on preferences, by expressing scepticism of the notion that preferences satisfy the rational deliberation and complete knowledge conditions that are embedded in the health state value elicitation instruments. However, Hausman rejects the view that the value of health is its impact on happiness, in part because the evidential connection between how good one feels and how well one's life is going in relation to one's health among other things is fragile. For example, adaptation to a poor state of health would imply that the health state does not impact considerably on happiness, and yet public health officials would still presumably want to improve the health of the relevant patient groups, a view that resembles Amartya Sen's contention that few would deliberatively choose to be a slave even if it was known that to be so would make one happy, because the opportunity to flourish would be substantively curtailed as a consequence (Sen, 1999). Indeed, as Hausman notes, some objectively poor health states, such as dementia, may sometimes directly improve happiness in the moment, and yet few would argue that this outweighs the severe limitations that these conditions impose. Conversely, but also because of the dissonance between mood and well-being, Hausman argues that people sometimes exaggerate the implications of often transitory health states on their longer term well-being, in particular in relation to acute bouts of anxiety and depression, when the perception of the bigger picture may be distorted. Thus, when the health state relates directly to mood, as in the case of anxiety and depression, a retrospective or prospective assessment may give a more accurate evaluation of the impact of that health state on the whole of one's life than if the utility of the health state was measured while it was running its course.

Hausman thus rejects Dolan and Kahneman's view that well-being equates to experienced utility or happiness. He does not believe that well-being consists in mental states and therefore considers it hopeless to measure the entirety of well-being with indicators of happiness, and he contends that the fact that the retrospective or prospective utility of an event often fails to equate with the utility experienced during that same event is not irrefutable evidence for discarding retrospective and prospective assessments. Indeed, remembered and decision utility, measured via retrospective and prospective assessments, might offer a reasonable indication of whether someone finds something worthwhile as opposed to enjoyable in the moment. Hausman leans towards retrospective methods as the most appropriate way in which to assess events, drawing on his own memories of reading to his two children, while barely remembering the tedium of taking care of them when they were sick or misbehaving. In short, Hausman implicitly aligns well-being to fulfilment, not to moment-by-moment happiness, and retrospective and prospective assessment may be tuned towards identifying the most fulfilling actions.

Hausman identifies a further concern that relates not to whether the value of an individual health state is its

impact on happiness, but to whether total well-being equates to a simple summation of the momentary or instant utilities of an event; i.e. whether instant utilities are additively separable. It is this concern that forms the focus of the rest of this essay. Specifically, Hausman (2015: 114) writes that as 'others have argued, a good life is not a sum of the net goodness of its moments . . . The same sum of momentary experiences can add up to a wonderful life or an incoherent and mediocre one, depending on how the experiences are ordered and what overall narrative they sustain'. Indeed, similar points have long been made in the history of thought. Hume (1983), for instance, across a range of enquiry, emphasized the importance of considering the whole rather than assessing the individual parts. Hausman, therefore, alludes to the distinction between maximizing experienced utility and being guided by remembered utility (as will be noted later, decision utility may often be subject to the same influences as remembered utility), a distinction that has received considerable attention in the psychology literature since the beginning of the 1990s but that has not yet infiltrated the health policy discourse in any substantive sense. The next section will elaborate a little on exactly how experienced utility differs from remembered and decision utility, a difference that until relatively recently was generally unknown.

Experienced Versus Remembered Utility

To recap, utility has had at least two distinct meanings in the history of economic thought. The first is 19th-century Benthamite utility, or classical utilitarianism, which prescribes that hedonic states ought to guide public policy, a return to which has been the preoccupation of the modern advocates of experienced utility maximization. By the end of the 19th century, most economists had come to think that it was impossible to derive an objective numerical index for the accurate measurement of interpersonal cardinal utility, and thus turned to decision utility, which forms the bedrock of modern welfare economics. They assumed that one could infer the differential utility of alternatives from choices, and that people will choose that which maximizes their experienced utility. To the extent that choices are often driven by prior experiences, it was, and still is, generally held that retrospective or remembered utility and decision utility will point in the direction of the maximization of experienced utility. From the early 1990s a body of evidence emerged that demonstrated that experienced utility often differs from remembered and decision utility (Kahneman *et al.*, 1997), basically because respondents frequently appear to focus upon salient aspects of temporally extended episodes when deciding and remembering, and do not integrate into the decision-making process all of the components of experienced utility, in particular the duration of an event. These salient aspects are often called gestalt characteristics (Ariely and Carmon, 2000), and include the tendency for respondents to place a heavy emphasis on the peak and end moments of an episode (so-called peak-end evaluation), a tendency for them to prefer worse outcomes to precede better outcomes rather than vice-versa, and an aversion to steep rates of change in outcomes (Ariely and Carmon, 2000). In short, echoing Hausman's concern, the evaluation of the whole often conflicts with the sum of the parts.

That individuals often tend to neglect the duration of events in their prospective and, in particular, retrospective evaluations can lead to violations of temporal monotonicity, a conflict with dominance that means that, in the assessment of two events, an individual remembers that which offered unambiguously greater experienced utility, or lower experienced disutility, as the worst of the two. Thus, the experiencing and the remembering selves sometimes differ, which may be necessary for people to function given the difficulties that persons suffering from hyperthymesia, that is, an autobiographical memory, encounter in their daily lives. That said, setting up the tension between the experiencing and remembering selves as an either–or conflict might not be the best course of action; both ought to be thought worthy of consideration in the policy discourse.

Of all the gestalt characteristics, peak-end evaluation, where retrospective assessment more closely correlates with an average of the best/worst and end moments of an experience than an integral of all the individual instant moments of that experience (and duration does not correlate at all), has attracted the most attention. There are several empirical tests that have demonstrated this phenomena (for a review, see Kahneman *et al.*, 1997), but perhaps the most famous, mentioned by Hausman himself, is that reported by Redelmeier *et al.* (2003), who divided 682 colonoscopy patients into two groups. In one of the groups, without informing the respondents, the colonoscope was left inserted for an additional minute at the end of the procedure, which would have caused some discomfort but, on average, at a reduced level than when the procedure was ongoing. The remembered disutility tended to be lower for this group, who on average experienced greater total disutility but had lower end moment discomfort than the

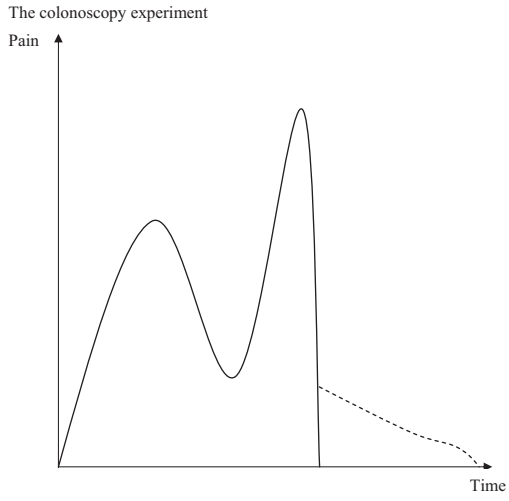

Figure 1. The colonoscopy experiment.

group for whom the colonoscope was immediately removed following the procedure. Figure 1 is a simple diagrammatic representation of what happened: the extra period of pain under the dashed line led to greater total pain when the solid and dashed lines are considered together, but is of lower intensity than the end point of the solid line, and thus the totality of that under the solid and dashed lines is remembered as being less unpleasant than when the event depicted by only the solid line is experienced.

Although not as extensive as the empirical literature that demonstrates a difference between experienced and remembered utility, there are a few studies that indicate that decision utility may also often be driven by the gestalts, a prominent example being the cold water experiment reported by Kahneman *et al.* (1993). In that study, participants were informed that they would be exposed to three trials, only two of which were conducted. In the first trial, the participants were required to immerse one of their hands in 14°C water for 60 s, an unpleasant experience. They were then allowed to dry their hand with a warm towel. In the second trial, a short time later, they were required to immerse their other hand in water for 90 s. For the first 60 s, the water temperature was again set at 14°C, but during the remaining 30 s the temperature was increased to a still unpleasant 15°C. For the third trial, which was not actually undertaken, the respondents were asked which of the two previous trials they would prefer to repeat; thus, they were being asked to make a decision. A significant majority expressed a preference for repeating the second trial, which imposed the greater total experienced disutility. The decisions of these respondents were driven by the relatively less unpleasant (but still absolutely unpleasant) end point of the second trial compared to the first trial, rather than the integrals of the moments of experienced instant utility in the two trials.

Peak-end evaluation offers important potential for personal and policy learning. For instance, if one wants to encourage patients to return for repeat medical procedures, there is a case for trying to make the end experience of each procedure as pleasant as possible. A similar point can be made with respect to healthy habit formation, where any individual might use end evaluation to trick himself or herself into further visits to the gym, for example. The behavioural scientist, Dan Ariely (2008), recounts an incident where end evaluation could have been put to good use in his own life. A serious burns victim in his youth, Ariely notes that the practice, common among nurses, of removing bandages as quickly as possible so as to limit the duration of pain (and distress to the nurses themselves) can cause pain so intense for the patient that it provokes serious mental trauma. Ariely advocates for a slower, more careful removal of bandages from burns patients in such circumstances, which may prolong the experience in the moment, but would help ensure that the peak and end moments of pain are less memorable.

Possible Implications for QALY Maximization

The gestalt literature that demonstrates that retrospective and prospective evaluations often give rise to remembered and decision utility that differ systematically from total experienced utility has tended to focus on short episodes or events. Nonetheless, in that the normative postulate of the most respected form of health economic evaluation among health economists—namely, cost-utility analysis (CUA)—is not too different from that of experienced utility maximization, it may reasonably be contended that the gestalt literature has some relevance when assessing temporally extended health profiles. CUA calls for the maximization of QALYs that, as with instant utility in experienced utility maximization, are assumed to be additively separable. The main difference between CUA and experienced utility maximization is that QALY weights are measures of health state utility elicited from respondents via hypothetical decision contexts (i.e. they are derived from

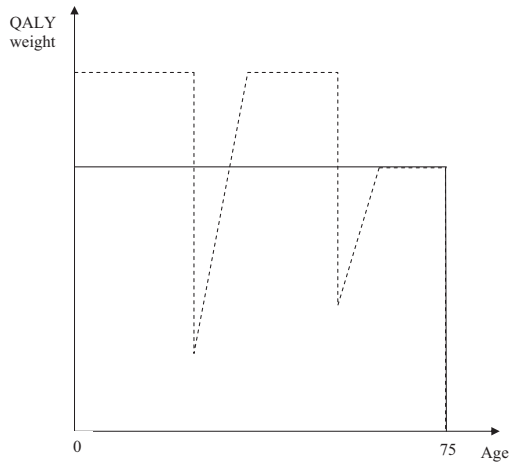

Figure 2. Peak/trough evaluation and QALY profiles.

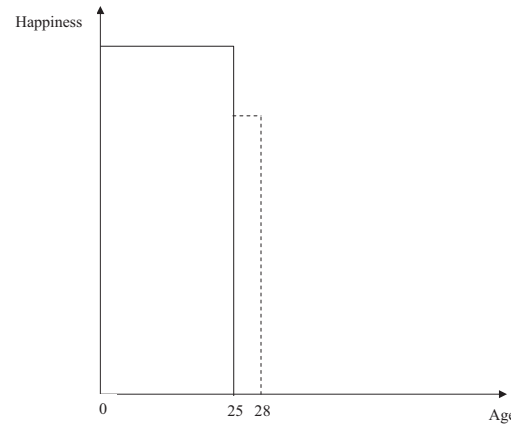

Figure 3. The James Dean effect.

methods that are informed by decision utility), and are not, therefore, experienced. It will by now be clear that the debate on which method (i.e. decision- or experienced-driven) might most satisfactorily be used to elicit indices of instant utility is not the focus of this essay. Given that CUA, similar to experienced utility maximization, assumes that policymakers ought to seek to maximize additively separable health-related utility when faced with a choice between competing temporally extended profiles, the relevant question here is that, in light of the gestalt characteristics, ought QALY maximization be the legitimate decision rule for any given patient or group of patients?

Oliver (2008) has reported a rare study in this area in which he tests for a variety of the gestalt characteristics by eliciting respondent choices between extended QALY profiles. He found considerable evidence of the effects of the gestalts on people's choices. For instance, in simple terms, Figure 2 illustrates a test he did of the peak/trough effect. The profile represented by the solid line is steady but comprises fewer lifetime QALYs than the profile given by the dashed line, which has a number of highs and lows. In a choice between these two profiles, a majority of Oliver's respondents preferred the stable profile, even though this offered fewer lifetime QALYs, a direct contradiction of the QALY maximization rule and an indication of the strong aversion that many people have to volatility and traumatic moments.

That some people may want to trade-off some total lifetime health to avoid substantial volatility in their health profile is not necessarily the consequence of an error in their decision-making. Indeed, there may be good reasons for why the choices of many align with peak-end evaluation. For instance, peaks and troughs give an indication of what is required to cope with a particular episode, and the end moment reveals the peak (Fredrickson, 2000). Echoing Hausman (and Hume), the remembering self may think that how a temporally extended experience fits together is normatively important, but this will be overlooked by aggregating moment-to-moment assessments of utility. Even if a normative justification cannot be found, the fact that many people attach disproportionate weight to, for example, an end moment in their retrospective assessments is a useful insight in informing personal and policy decisions, as discussed above (for instance, even if one does not think that people ought to disproportionately weight the end moment of an experience, the fact that they often do might help policymakers to design interventions that more effectively meet their policy objectives, such as better motivating people to undertake necessary repeat medical examinations). However, focusing on the normative, it might reasonably be contended that we perhaps ought to be circumspect of some of the evidence reported in the gestalt literature, with the so-called James Dean effect being a case in point.

The James Dean effect is the observation reported by Diener *et al.* (2001) that many respondents rate a wonderful life that ends abruptly as better than one that is identical apart from having additional less wonderful but still worthwhile years, a further violation of temporal monotonicity that appears also to be motivated by end evaluation. Figure 3 diagrammatically depicts the phenomena: the profile under the solid line that results in an abrupt death at the age of 25 years in this case will often be preferred over an identical life with an additional, say, 3 years of less enjoyable life at the end.

Oliver (2008) observed similar results with respect to QALY profiles, but one can reasonably question whether these responses reflect legitimate preferences, or whether they are rather the result of an encoding heuristic that causes respondents to be influenced superficially rather than prompting a deliberative consideration of everything that the profiles imply (Ariely and Carmon, 2000). Given the human propensity to cling on to life, deliberation might suggest that most people aged, say, 23 or 24 years, would, in the context of Figure 3, very much want to live on to an age of 28 years, even if happiness diminished (but remained positive) in the final 3 years, if the only alternative is an abrupt death at the age of 25 years.

Conclusion

From the relevant literature, it appears that retrospective and prospective evaluations, and hence remembered and decision utility, are often influenced by the gestalt characteristics, and that these pose some difficult questions for policymakers in deciding the best course of action when faced with competing, temporally extended profiles. That is, should the gestalt characteristics and their consequences such as duration neglect and violations of temporal monotonicity be respected normatively?

For some, violations of experienced utility maximization are problematic. From this perspective, significant duration neglect is an error, at least in part brought on by encoding and the erroneous human tendency to remember only salient moments of past experiences rather than all negative and positive contributions to their lifetime happiness. Advocates of experienced utility maximization call for a method by which utility is assessed as it is experienced by respondents to give a more accurate picture of how much benefit particular events generate. It can reasonably be argued, however, that particular moments have a long-term effect on respondent well-being that is not captured in the experienced instances of the event itself, and that it is impractical to capture these benefits in memories experienced in future moments. In some domains with long-term effects, such as the assessment of the benefits promised by new healthcare interventions, the use of experienced utility measurements is also impractical: by the time the experienced utility of interventions with distant consequences is assessed, the intervention might be obsolete. Moreover, as argued by Hausman, for many an integration of instant moments of happiness may simply be a misguided way in which to assess the well-being generated by the important events in their lives; they may be more interested in how fulfilling these events are or have been in relation to the overall narrative of their lives, which might reasonably be assessed via retrospective and prospective evaluations.

Although an encoding heuristic may have plausibly driven some of the findings uncovered in the gestalt literature, and although the utility that people experience presumably ought to be at least a partial consideration for responsible policymakers, the advocates of experienced utility maximization are perhaps too confident in their policy prescriptions. It does not seem unreasonable, for example, for a person to want to deliberatively and deliberately trade-off some total experienced utility or health so as to try to avoid volatility or heavy troughs in his/her lifetime profile. If such observations are common, as they appear to be, then a case can be made for including such considerations in policy deliberations. The import of the gestalt characteristics is thus likely to be contextual, and, as implied by Hausman, the answer to the question of whether they ought to be respected normatively is therefore probably yes, sometimes.

References

Ariely, D. (2008). *Predictably Irrational: The Hidden Forces that Shape our Decisions*. London: HarperCollins.

Ariely, D. and Carmon, Z. (2000). Gestalt Characteristics of Experiences: The Defining Features of Summarized Events. *Journal of Behavioral Decision Making*, **13**, 191–201.

Diener, E., Wirtz, D., and Oishi, S. (2001). End Effects of Rated Life Quality: The James Dean Effect. *Psychological Science*, **12**, 124–128.

Dolan, P. and Kahneman, D. (2008). Interpretations of Utility and Their Implications for the Valuation of Health. *Economic Journal*, **118**, 215–234.

Drummond, M. F., Sculpher, M. J., Torrance, G. W., O'Brien, B. J., and Stoddart, G. L. (2005). *Methods for the Economic Evaluation of Health Care Programmes*, 3rd edn. Oxford: Oxford University Press.

Fredrickson, B. L. (2000). Extracting Meaning from Past Affective Experiences: The Importance of Peaks, Ends, and Specific Emotions. *Cognition and Emotion*, **14**, 577–606.

Hausman, D. M. (2015). *Valuing Health: Well-Being, Freedom and Suffering*. Oxford: Oxford University Press.

Hume, D. (1983). *An Enquiry Concerning the Principles of Morals*. Indianapolis: Hackett Publishing Company.

Kahneman, D., Wakker, P. P., and Sarin, R. (1997). Back to Bentham? Explorations of Expected Utility. *The Quarterly Journal of Economics*, **112**, 375–405.

Kahneman, D., Fredrickson, B. L., Schreiber, C. A., and Redelmeier, D. A. (1993). When More Pain is Preferred to Less: Adding a Better end. *Psychological Science*, **4**, 401–405.

Oliver, A. (2008). Assessing the Influence of Gestalt-type Characteristics on Preferences Over Lifetime Health Profiles. *Medical Decision Making*, **28**: 723–731.

Redelmeier, D., Katz, J., and Kahneman, D. (2003). Memories of Colonoscopy: A Randomized Trial. *Pain*, **104**, 187–194.

Sen, A. (1999). *Development as Freedom*. New York: Random House.