New Challenges to the Rationality Assumption
Author(s): Daniel Kahneman
Published by: Mohr Siebeck GmbH & Co. KG
Stable URL: http://www.jstor.org/stable/40753012
Accessed: 17-01-2018 14:09 UTC

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at http://about.jstor.org/terms

Mohr Siebeck GmbH & Co. KG is collaborating with JSTOR to digitize, preserve and extend access to Journal of Institutional and Theoretical Economics (JITE) / Zeitschrift für die gesamte Staatswissenschaft
New Challenges to the Rationality Assumption

by

DANIEL KAHNEMAN *

Abstract

In contrast to logical criteria of rationality, which can be assessed entirely by reference to the system of preferences, substantive criteria of rational choice refer to an independent evaluation of the outcomes of decisions. One of these substantive criteria is the experienced hedonic utility of outcomes. Research indicates that people are myopic in their decisions, may lack skill in predicting their future tastes, and can be led to erroneous choices by fallible memory and incorrect evaluation of past experiences. Theoretical and practical implications of these challenges to the assumption of economic rationality are discussed. (JEL: A 00)

1. Introduction

The assumption that agents are rational is central to much theory in the social sciences. Its role is particularly obvious in economic analysis, where it supports the useful corollary that no significant opportunity will remain unexploited. In the domain of social policy, the rationality assumption supports the position that it is unnecessary to protect people against the consequences of their choices. The status of this assumption is therefore a matter of considerable interest. This article will argue for an enriched definition of rationality that considers the actual outcomes of decisions, and will present evidence that challenges the rationality assumption in new ways.

The criteria for using the terms “rational” or “irrational” in non-technical discourse are substantive: one asks whether beliefs are grossly out of kilter with available evidence, and whether decisions serve or damage the agent’s interests. In sharp contrast, technical discussions of rationality generally adopt a logical

* Presented at the 11th International Seminar on the New Institutional Economics, Wallerfangen/Saar, Germany, June 1993, and at the IEA Conference on Rationality and Economics, Turin, Italy, October, 1993. A different version was presented at a Political and Economic Analysis Workshop in Honor of P. Zusman, in Rehovot, Israel, June 1993. The research leading to this paper was supported by grants from Sloan Foundation, the McArthur Foundation and the National Science Foundation. I am greatly indebted to Amos Tversky for many discussions of the issue of rationality over the years, and for insightful comments on drafts of this paper. He should not be assumed to agree with all I say. Alan Schwartz provided helpful editorial assistance.
conception, in which an individual's beliefs and preferences are said to be rational if they obey a set of formal rules such as complementarity of probabilities, the sure thing principle or independence of irrelevant alternatives. In the laissez-faire spirit of modern economics and decision theory, the content of beliefs and of preferences is not a criterion of rationality – only internal coherence matters (Sen [1993]). The methodology of the debate reflects this concern for consistency: in the classic paradoxes of Allais and Ellsberg, for example, two intuitively compelling preferences are shown to be jointly incompatible with the axioms of expected utility theory, though each preference is unobjectionable on its own. Irrational preferences are diagnosed without having to observe anything that is not a preference.

Some authors have been dissatisfied with the exclusive focus on consistency as a criterion of rationality. Thus, Sen [1990, 210] has written: "Rationality may be seen as demanding something other than just consistency of choices from different subsets. It must, at least, demand cogent relations between aims and objectives actually entertained by the person and the choices that the person makes. This problem is not eliminated by the terminological procedure of describing the cardinal representation of choices as the "utility" of the person, since this does not give any independent evidence on what the person is aiming to do or trying to achieve." This article will ask whether there exists a cogent relation between a person's choices and the hedonic consequences of these choices.

In spite of occasional attempts to broaden the scope of the rationality debate in decision theory, the patterns of preference discovered by Allais and Ellsberg have been at the center of this debate for several decades. It is often implied that if these paradoxes can be resolved, then economic analysis can safely continue to assume that agents are rational. The focus on paradoxes has indirectly strengthened the rationality dogma: if subtle inconsistencies are the worst indictment of human rationality, there is indeed little to worry about. Furthermore, the preferences that Allais and Ellsberg described do not appear foolish or unreasonable, and lay people as well as many theorists believe they can be defended (Slovic and Tversky [1974]). Indeed, the ambiguous normative status of the Allais and Ellsberg patterns has inspired many attempts to reconcile observed preferences with rationality by adopting a more permissive definition of rational choice (Tversky and Kahneman [1986]).

More recent challenges to the rationality assumption do not lend themselves to such attempts at reconciliation. Numerous experiments illustrate beliefs and preferences that violate a fundamental requirement variously labeled extensionality (Arrow [1982]), consequentialism (Hammond [1985]) or invariance (Tversky and Kahneman [1986]). The same choice problem may evoke different preferences, depending on inconsequential variations in the formulation of options (Tversky and Kahneman [1986]) or in the procedure used to elicit choices (Tversky, Slovic and Kahneman [1990]). The main method of this research still involves the documentation of pairs of preferences, each accept-
able on its own, which jointly violate an axiom of invariance. These inconsist-
encies are more difficult to rationalize than the classic paradoxes, because invari-
ance is a more compelling axiom of rational choice than cancellation or inde-
pendence (Tversky and Kahneman [1986]). Some examples of this research
will be presented below.

The present treatment attempts to supplement the logical analysis of prefer-
ences by introducing substantive criteria of rationality. Unlike the logical anal-
ysis, a substantive criterion is external to the system of preferences. It requires
some way of assessing outcomes as they occur, not only as they are conceived
at the time of decision. The substantive question on which we focus here is
whether choices maximize the (expected) utility of their consequences, as these
consequence will actually be experienced. Accurate prediction of future tastes
and accurate evaluation of past experiences emerge as critical elements of an
individual's ability to maximize the experienced quality of his outcomes.
Demonstrated deficiencies in the ability to predict future experiences and to
learn from the past emerge as new challenges to the assumption of rationality.
More provocatively, the observed deficiencies suggest the outline of a case in
favor of some paternalistic interventions, when it is plausible that the state
knows more about an individual's future tastes than the individual knows
presently. The basis of these developments is an analysis of the concept of
utility, which is introduced in the next section.

2. Multiple Notions of Utility

The term "utility" can be anchored either in the hedonic experience of out-
comes, or in the preference or desire for that outcome. In Jeremy Bentham's
usage, the utility of an object was ultimately defined in hedonic terms, by the
pleasure that it produces. Others have interpreted utility as "wantability"
(Fisher [1918]). Of course, the two definitions have the same extension if people
generally want that which they will eventually enjoy – a common assumption
in discussions of utility. Economic analysis is more congenial to wants and
preferences than to hedonic experiences, and the current meaning of utility in
economics and decision research is a positivistic version of wantability: utility
is a theoretical construct inferred from observed choices. This definition has
been thoroughly cleansed of any association with hedonistic psychology, and
of any reference to subjective states.

The present analysis starts with two observations. The first is that the
methodological strictures against a hedonic notion of utility are a relic of an
earlier period in which a behavioristic philosophy of science held sway. Subject-
ive states are now a legitimate topic of study, and hedonic experiences such as
pleasure, pain, satisfaction or discomfort are considered open to useful forms
of measurement. The second observation is that it may be rash to assume as a
general rule that people will later enjoy what they want now. The relation
between preferences and hedonic consequences is better studied than postulated.

These considerations suggest an explicit distinction between two notions of utility. The *experienced utility* of an outcome is the measure of the hedonic experience of that outcome. This is similar to Bentham’s awkward use; the first footnote of his book was properly apologetic about the poor fit of the word “utility” to pleasure and pain, but he found no better alternative. The *decision utility* of an outcome, as in modern usage, is the weight assigned to that outcome in a decision.

The distinction between experienced utility and decision utility opens new avenues for the study of rationality. In addition to the syntactic criterion of consistency, we can now hope to develop a substantive/hedonic criterion for the rationality of a decision: does it maximize the expectation of experienced utility? Of course, this criterion is not exhaustive, and its adoption implies no commitment to a hedonistic philosophy. As Sen has often pointed out (e.g., Sen [1987]), the maximization of (experienced) utility is not always “what people are trying to achieve.” It is surely the case, however, that people sometimes do try to maximize pleasure and minimize pain, and it may be instructive to drop the assumption that they perform this optimization task flawlessly.

Errors in the assignment of decision utility to anticipated outcomes can arise from inaccurate forecasting of future hedonic experience. Correct prediction of future tastes is therefore one of the requirements of rational decision making (March [1978]). Kahneman and Snell [1990] defined the *predicted utility* of an outcome as the individual’s beliefs about its experienced utility at some future time. Two sets of empirical questions arise: (i) How much do people know about their future tastes? Is it likely that an objective observer (or a government) could make more accurate predictions than individuals would make on their own behalf? (ii) Do people adequately consider the uncertainty of their future tastes in making decisions? Are decision utilities adequately informed by reasoned beliefs about experienced utility?

Additional issues arise because of possible disparities between memory and actual hedonic experience. Outcomes are commonly extended over time, and global evaluations of such outcomes are necessarily retrospective – and therefore subject to errors. Examples of substantial discrepancies between *retrospective utility* and *real-time utility* are discussed below.

The restoration of Bentham’s notion of utility as an object of study evidently sets a large agenda for theoretical and empirical investigation. The following sections summarize highlights of what has been learned in early explorations of this agenda. Decision utility, predicted utility, and the relations between real-time and retrospective utility are discussed in turn. The final section reviews possible implications of the findings for the rationality debate.
3. Some Characteristics of Decision Utility

Decision utility has long been a topic of study, and much is known about it. The following discussion selectively addresses three research conclusions that are of particular relevance to the issue of rationality, as it is construed in this paper. (i) Carriers of utility. The main carriers of decision utility are events, not states; in particular, utility is assigned to gains or losses relative to a reference point which is often the status quo (Kahneman and Tversky [1979]). (ii) Loss aversion. Losses loom larger than corresponding gains (Kahneman and Tversky [1979]; Tversky and Kahneman [1991]). (iii) Framing effects. The same objective outcomes can be evaluated as gains or as losses, depending on the framing of the reference state (Tversky and Kahneman [1986]).

An early observation that illustrates points (i) and (iii) above was labeled the isolation effect (Tversky and Kahneman [1986]).

Problem 1. Assume yourself richer by $300 than you are today. You have to choose between
a sure gain of $100
50% chance to gain $200 and 50% chance to gain nothing

Problem 2. Assume yourself richer by $500 than you are today. You have to choose between
a sure loss of $100
50% chance to lose nothing and 50% chance to lose $200.

It is easily seen that the two problems are extensionally equivalent in terms of wealth: both offer a choice between a state in which wealth is increased by $400 and a gamble with equal chances to increase current wealth by $300 or by $500. If people spontaneously evaluate options in these terms they will choose the same option in the two problems − but observed preferences favor the sure thing in Problem 1 and the gamble in Problem 2. Because the equivalence of the two problems is intuitively compelling when it is pointed out, the difference between the responses they elicit is a framing effect: an inconsequential feature of the formulation strongly affects preferences. Most important in the present context, the experiment demonstrates that people are content to assign utilities to outcomes stated as gains and losses, contrary to the standard assumption that the carriers of utility are states of wealth.

Figure 1 exhibits loss aversion in a schematic value function: the function is steeper in the domain of losses than in the domain of gains. The ratio of the slopes in the two domains, called the loss aversion coefficient, has been estimated as about 2 : 1 in several experiments involving both risky and riskless options (Tversky and Kahneman [1991], [1992]). Figure 2 (from Kahneman, Knetsch and Thaler [1991]) illustrates the role of a reference point in the evaluation of a transaction. The choice illustrated in this figure is between a state (Point A) with more of Good Y and a state (Point D) with more of Good X. The hy-
Hypotheses about the carriers of utility and about framing effects entail that the preference between A and D could differ depending on the current reference point – contrary to a substantial body of economic theory. Consider the choice between A and D from C. This is a positive choice between two gains, in
Good X or in Good Y. If the reference is A, however, the two options are framed quite differently. One possibility is to retain the status quo by staying at A. The alternative is to accept a trade that involves the conjunction of a loss in Good Y and a gain in Good X. The C-A interval is evaluated as a gain in the first frame (from C), but the same interval is evaluated as a loss from A. Because of loss aversion, the impact of the C-A difference is expected to be greater in the latter case than in the former. We therefore predict a systematic difference between the preferences in the two frames: if people are indifferent between A and D from C, they should strictly prefer A over D from A (Tversky and Kahneman [1991]).

The predicted result, known as the endowment effect (Thaler [1980]), has been confirmed in several laboratories. Subjects in one condition of an experiment by Kahneman, Knetsch and Thaler [1990] were offered a choice between a decorated mug (worth about $6 at the University bookstore) and a sum of money; they answered a series of questions to determine the amount of money at which they were indifferent between the two options. Other subjects were first given a mug; they answered similar questions to indicate the amount of money for which they would just agree to exchange it. The subjects had no strategic incentive to conceal their true values. A critical feature of the study is that the choosers in the first group and the mug owners in the second group faced objectively identical options: they could leave the experimental situation owning a new mug, or with extra money in their pocket. The analysis of figure 2 applies, however. If the mug is Good Y, the C-A interval (the difference between having a mug and not having one) is evaluated as a gain by the choosers, and as a loss by the mug owners. As predicted, the average cash value of a mug was much larger for the owners ($7.12 in one experiment) than for choosers ($3.50). A significant (but somewhat smaller) difference between owners and choosers was observed in a replication by Franciosi, Kujal, Mischelitsch and Smith [1993].

Implications of the endowment effect for various aspects of economic and legal theory have been discussed extensively elsewhere (Ellickson [1989]; Hovenkamp [1991]; Hardie, Johnson and Fader [1993]; Kahneman, Knetsch and Thaler [1991]). The effect is relevant to the present treatment because it implies that decision utilities may be extremely myopic. The subjects in the mugs experiment made a decision that was to have consequences over the relatively long term: a coffee mug is an object that one may use daily, sometimes for years. The long-term states between which the subjects had to choose — "own this mug" or "not own this mug" — were the same for all subjects. The large difference between the preferences of owners and choosers indicates that these enduring states were not the main object of evaluation. The effective carriers of utility were the transitions that distinguished the experimental treatments: "receive a mug" or "give up your mug." In this experiment, and perhaps in many other situations, people who make decisions about a long-term state appear to use their evaluation of the transition to that state as a proxy.
The results of the mugs experiment present two overlapping challenges to the assumption of rationality. The logical notion of rationality is violated by the inconsistent preferences observed in different representations of the choice between a mug and money. A substantive condition of rationality is violated if the endowment effect is viewed as a costly manifestation of extreme myopia. An agent who routinely uses transient emotions as a proxy for the utility of long-term states is manifestly handicapped in the achievement of good outcomes.

4. Predicted Utility: Do People Know What They Will Like?

Although the constancy of underlying tastes is a matter of theoretical debate, the following proposition will not be controversial: the hedonic experience associated with a particular stimulus or with a particular act of consumption is susceptible to large changes over time and over varying circumstances. Some cyclical changes of experienced utility are regular and readily predictable: ingesting the same food may evoke delight in a state of hunger, disgust in satiation. At the other extreme, radical changes of circumstances produce adaptations and changes of experienced utility that violate common expectations. A well-known psychological study showed that most paraplegics adapt far better than most people would predict, and that lottery winners are generally less happy in the long run than the common fascination with lotteries might suggest (Brickman, Coates and Janoff-Bulman [1978]).

Many decisions explicitly or implicitly involve predictions of future consumption and of future utility (March [1978]). An encyclopaedia may not be worth buying if one will not use it, the premium paid for a house with a view may be wasted if the view ceases giving pleasure after a time, and a medical procedure that improves survival chances should perhaps be rejected by a patient who is likely to find life without vocal cords intolerable.

How accurately do people predict their future utility? Most of the evidence about this question is indirect. Thus, it is suggestive that some important results of hedonic research are generally considered counter-intuitive (Kahneman and Snell [1990]). The surprises include the striking increase of liking by mere exposure to initially neutral stimuli, and some effects of dissonance on tastes. A study of people's intuitions about possible ways to induce a child to like or to dislike a food showed a similar lack of collective wisdom about the dynamics of taste. Dynamic inconsistency may be another manifestation of inaccurate hedonic prediction. For example, Christensen-Szalanski [1984] documented the incidence of cases in which women in labor reversed a long-standing preference for delivery without anaesthetics. The reversals could be due to improper discounting of the pain in the initial preferences; they could also reflect an error in the initial prediction of the intensity of labor pains.

Loewenstein and Adler [1993] observed a remarkable result in a study of the endowment effect. They showed subjects a mug engraved with a decorative
logo, and asked some of these subjects to "... imagine that we gave you a mug exactly like the one you can see, and that we gave you the opportunity to keep it or trade it for some money." The subjects then filled out a form indicating their preferences for a range of stated prices, following the procedure of Kahneman, Knetsch and Thaler [1990]. The mean predicted selling price was $3.73. Next, all subjects were given a mug and a second form, which actually provided an opportunity to exchange the mug for cash. The mean selling price for the subjects who had made a prediction a few minutes earlier was $4.89, significantly higher than the predicted value, and only moderately lower than the selling price of $5.62 stated by subjects who had not made a prediction. The subjects in this experiment were apparently unable to anticipate that possession of the mug would induce a reluctance to give it up.

Simonson [1990] reported a result that illustrates a failure of hedonic prediction – or perhaps a failure to make such a prediction. Simonson gave students an opportunity to select from a choice set of snacks at the beginning of a class meeting; they received their selections at the end of the class. Subjects in one experimental condition made one choice each week for three weeks. In another condition subjects made choices for all three weeks at the first session. The choices made by the two groups were strikingly different. Subjects who chose a snack on three separate occasions tended to choose the same snack or a closely similar one every time. In contrast, subjects who chose in advance for three weeks tended to pick different items for the different occasions. It is reasonable to view these variety-seeking choices as erroneous: the subjects apparently failed to realize that their current preferences would be restored after a one-week interval. A further study clarified the nature of the error. Anticipatory choices were less variable when subjects were asked, before indicating a decision, to predict the preferences they would actually have on the subsequent occasions of testing. This finding suggests that the subjects were in fact able to predict their future preferences accurately. In the absence of a special instruction, however, they did not take the trouble to generate a prediction of their future taste before making a decision about future consumption.

Kahneman and Snell [1992] reported an exploratory study of the accuracy of hedonic prediction. They examined predictions of future liking for a food item or a musical piece, under conditions that made a change of attitude likely. In an initial experiment the subjects consumed a helping of their favorite ice cream while listening to a particular piece of music, at the same hour on eight consecutive working days under identical physical conditions. Immediately after each episode they rated how much they had liked the ice cream and the music. At the end of the first session they predicted the ratings they would make on the following day, and on the final day of the experiment. This experiment was intended to test the accuracy of hedonic predictions under relatively favorable conditions. We reasoned that student subjects have not only had much experience consuming ice cream and listening to music; they have had experience with repeated consumption of these items, and could therefore be expected
to anticipate the effect of frequent repetition on their tastes. Other experiments in the series used a stimulus that is less familiar and less popular than ice cream in the student population – plain low-fat yogurt.

The accuracy of hedonic predictions was generally quite poor. A comparison of the average of predictions to the average of the actual ratings revealed some shared failures to anticipate common trends in the hedonic responses. For example, most subjects predicted, after tasting one spoonful of plain low-fat yogurt, that they would assign the same rating to a 6 oz helping on the next day. In fact, the larger helping was a much worse experience. Most subjects also failed to anticipate the considerable improvement in the attitude to plain yogurt which occurred (for most of them) with further exposure to that substance. There apparently exists a lay theory of hedonic changes, of mediocre accuracy, which most of our subjects accepted. Another analysis was concerned with individual differences in predictions and in actual hedonic changes. There was substantial variability in both measures, but the correlation between them was consistently close to zero. The data provided no indication that individuals were able to predict the development of their tastes more accurately than they could predict the hedonic changes of a randomly selected stranger.

The results of these studies suggest two conclusions. (i) People may have little ability to forecast changes in their hedonic responses to stimuli (Kahneman and Snell [1992]; Loewenstein and Adler [1993]). (ii) Even in situations that permit accurate hedonic predictions, people may tend to make decisions about future consumption without due consideration of possible changes in their tastes (Simonson [1990]). If supported by further research, these hypotheses about the accuracy of predicted utility and about its impact on decision utility would present a significant substantive challenge to the assumption of rationality.

The properties of predicted utility have implication for other domains. Consider the issue of informed consent to an operation that will change the patient’s life in some significant way. The normal procedure for consent emphasizes the provision of objective information about the effects of surgery. However, truly informed consent is only possible if patients have a reasonable conception of expected long-term developments in their hedonic responses, and if they assign appropriate weight to these expectations in the decision. A more controversial issue arises if we admit that an outsider can sometimes predict an individual’s future utility far better than the individual can. Does this superior knowledge carry a warrant, or even a duty, for paternalistic intervention? It appears right for Ulysses’ sailors to tie him to the mast against his will, if they believe that he is deluded about his ability to resist the fatal call of the sirens.

5. Real-Time and Retrospective Utility – Do People Know What They Have Liked?

Retrospective evaluations of the experienced utility of past episodes are undoubtedly the most important source of predictions of the hedonic quality of
future outcomes. The experiences of life leave their traces in a rich store of
evaluative memories, which is consulted, apparently automatically, whenever a
significant object or experience is brought to mind (Zajonc [1980]). The system
of affective and evaluative memories may be independent of any ability to recall
the incidents that produced an attitude. Thus, people often recognize that they
like or dislike a person they have met before, without knowing why. Evaluative
memories are immensely important because they contain the individual's accu-
mulated knowledge of stimuli that are to be approached and of others that are
to be avoided. Indeed, the only form of utility that people could possibly learn
to maximize is the anticipated utility of future memories. Every individual has
the lifelong habit of trusting memories of past episodes to guide choices among
future outcomes. As we shall see, however, trusted evaluative memories are
sometimes deceptive.

Although retrospective evaluations and affective memories define what
is learned from the past, they are not the ultimate criterion of experienced
utility. Hedonic or affective quality is an attribute of each moment of experi-
ence; the sign and intensity of the experience may vary considerably even over
the course of a brief episode, such as drinking a glass of wine. The retro-
spective evaluation of an extended episode necessarily involves two operations:
the recollection of the momentary experiences that constituted the episode, and
an operation that combines the affect of these moments into a global evalua-
tion. Because both operations are fallible, retrospective evaluations should be
viewed with greater distrust than introspective reports of current experience.
The effects of defective memory are sometimes painfully obvious: people who
care for an elderly parent often observe that they accept their parent's immedi-
ate responses to the current situation with normal respect, even as they dismiss
most retrospective evaluations as unreliable. The difficulties that arise in sum-
marizing an episode by a global evaluation are more subtle, but no less signif-
icant.

There are strong normative intuitions about the correct way to combine the
utilities of a continuous series of experiences into a global evaluation. A prin-
ciple of temporal integration has considerable appeal: the utility of an episode
extended over time is the integral of momentary hedonic value over the dura-
tion of the episode. The justification for temporal integration is the assumption
that successive selves should be treated equally, an assumption so compelling
that a general case for utilitarianism has been built on it (Parfit [1984]). Even
more appealing than temporal integration is the principle of temporal
monotonicity. Consider two episodes that are preceded and followed by a steady
state of hedonic neutrality. Assume that the second episode is obtained by
adding an unanticipated period of pain (or pleasure) to the first, prior to the
return to the neutral state. The monotonicity principle asserts that the hedonic
quality of the added period determines whether the longer episode has higher
or lower global utility than the shorter. In other words, adding pain at the
end of an episode must make it worse; adding pleasure must make it better.¹

Several recent studies indicate that retrospective evaluations obey neither temporal integration nor temporal monotonicity. The studies conducted so far have dealt with episodes that were brief and uniform, both in content and in the sign of the hedonic experience, either non-negative or non-positive throughout. Several experiments involved controlled exposure to affect-inducing stimuli (films of pleasant or unpleasant content; loud unpleasant sounds; immersion of a hand in painfully cold water). Subjects used an “affect meter” to provide a continuous record of their momentary hedonic response during some of these episodes. Later they also provided retrospective global evaluations of the “overall discomfort” or “overall pleasure” of the episodes, and in some cases chose an episode to which they would be exposed again. In one non-experimental study (Reidelmeier and Kahneman [1993]) patients undergoing a colonoscopy for medical reasons provided reports of pain every 60 seconds, as well as subsequent global evaluations and measures of preference for the entire episode.

The results of these studies support two empirical generalizations. (1) “The Peak & End Rule”: global evaluations are predicted with high accuracy by a weighted combination of the most extreme affect recorded during the episode and of the affect recorded during the terminal moments of the episode. Here again, as in the context of decision utility, the evaluation of particular moments appears to be used as a proxy for the evaluation of a more extended period of time. (2) Duration Neglect. The retrospective evaluation of overall or total pain (or pleasure) is not independently affected by the duration of the episode. In the colonoscopy study, for example, the duration of the procedure varied from 4 to 69 minutes in a sample of 101 patients. Surprisingly, these variations of duration had no significant effect on retrospective evaluations. The ratings of both patients and attending physicians were dominated by the intensity of pain at its worst, and by the intensity of discomfort during the last few minutes of the procedure. Duration neglect is not immutable, of course: people can judge the duration of episodes with fair accuracy, and will treat this attribute as relevant when their attention is explicitly drawn to it (Varey and Kahneman [1992]). In general, however, affective peaks and endings are more salient than duration in the cognitive representation of events.

Figure 3 is taken from a study that examined violations of the rule of temporal monotonicity in a choice between painful episodes (Kahneman et al. [1993]). Paid volunteers expected to undergo three experiences of moderate physical

¹ The temporal monotonicity principle does not apply if the addition of pain or pleasure to the episode alters hedonic after-effects, such as relief, after-glow, or the affect associated with subsequent recollection. More generally, the analysis of experienced utility becomes difficult to apply where the consumption of memories plays an important role (Elster and Loewenstein [1992]).
pain during an experimental session. In fact they only had two trials. In the Short trial the subject held one hand in water at 14°C for 60 seconds, then immediately dried his hand with a towel. In the Long trial, the subject held the other hand in water for a total of 90 seconds. During the first 60 seconds of the Long trial the temperature of the water was 14°C, just as in the Short trial; during the extra 30 seconds the temperature of the water was gradually raised to 15°C, still unpleasant but for most subjects a clear improvement over the initial state. The order of the two trials was varied for different subjects. A few minutes after the second trial, the subjects were reminded that they were due to have another trial and were asked which of the two preceding experiences they chose to repeat.

The curves shown in figure 3 present average momentary ratings of discomfort for the Long trial, separately for two groups of subjects who showed different patterns of response: the majority who indicated decreasing discomfort as the temperature of the water was raised, and a minority who reported little change of discomfort. The choices of which trial to repeat were markedly different in these two groups: 17 of the 21 subjects whose discomfort diminished preferred to repeat the Long trial, in violation of temporal monotonicity; only 5 of the 11 subjects whose discomfort did not change preferred the Long trial. The results of both groups conform to the Peak & End rule and exhibit duration neglect. For the minority whose pain did not diminish, the peak and the end of the pain were at the same level (see figure 3), and were the same in
the Short and in the Long trials. The Peak & End rule predicts that these subjects should evaluate the two trials alike, a prediction that is confirmed by the nearly even split of preferences. For the larger group of subjects whose pain diminished at the end of the Long trial, the Peak & End rule predicts that this trial should be less aversive than the Short one, and the choice data again confirm the prediction. Overall, about 2/3 of subjects violate dominance in this situation, a robust result that has been replicated with numerous subjects under slightly different conditions.

Additional analyses clarify the mechanism that produces these violations of temporal monotonicity: most subjects erroneously believed that the coldest temperature to which they had been exposed was not the same in the two trials: their memory of the worst moment of the Long trial was mitigated by the subsequent improvement. Our evidence suggests that episodes are evaluated by a few “snapshots” rather than by a continuous film-like representation (Fredrickson and Kahneman [1993]). The snapshots are in fact montages, which may blend impressions of different parts of the experience. The overall experience is judged by a weighted average of the utility of these synthetic moments.

Other experiments showed that subjects who are only given verbal descriptions of the trials generally prefer the Short one, in accordance with the principle of temporal monotonicity. Telling these subjects that their memory of the Long trial will be more favorable does not diminish their preference for the Short trial. This observation indicates that the participants in the original experiment did not deliberately apply a policy of selecting the experience that would leave them with the most pleasant memory. However, subjects who have had personal experience of the two trials are quite reluctant to abandon their preference for the Long trial even when the nature of the two trials is carefully explained after the fact. It is evidently not easy to overcome a lifetime habit of trusting one’s evaluations of personal memories as a guide to choice.

The studies reviewed in this section have documented a consistent pattern of violations of a compelling normative rule. The axiom of temporal monotonicity is a substantive principle of rationality, a variant of “more is better” formulations of dominance. The violations of this principle have been traced to basic cognitive processes that produce representations and evaluations of episodes. The requirements of substantive rationality are apparently not compatible with the psychology of memory and choice.

The results of the cold-water study illustrate an ethical dilemma that was extensively discussed by Schelling [1984]. The history of an individual through time can be described as a succession of separate selves, which may have incompatible preferences, and may make decisions that affect subsequent selves. In the cold-water experiment, for example, the experiencing subject who records momentary affect and the remembering subject who makes retrospective evaluations appear to have conflicting evaluations. Which of these selves should be granted authority over outcomes that will be experienced in the
future? The principle of temporal monotonicity assigns priority to the experiencing subject. In the normal conduct of life, however, it is the remembering subject who assumes the all-important role of laying out guidelines for future actions. Is there an ethical justification for favoring one of these evaluations over the other? This question has immediate implications for the application of rules of informed consent in medical practice. Imagine a painful medical procedure that lasts a specified number of minutes and ends abruptly with pain at its peak. We have seen that the physician could probably ensure that the patient will retain a more favorable memory of the procedure by adding to it a medically superfluous period of diminishing pain. Of course, the patient would probably reject the physician’s offer to provide an improved memory at the cost of more actual pain. Should the physician go ahead anyway, on behalf of the patient’s future remembering self? This dilemma illustrates a class of problems of paternalism that are likely to arise in many policy debates, if considerations of experienced utility are assigned the weight they deserve in these debates.

6. General Discussion

The standard theory of choice provides a set of conditions for rationality that may be necessary, but are hardly sufficient: they allow many foolish decisions to be called rational. This essay has argued that it is generally useful and sometimes possible to supplement the logical analysis of decisions by substantive criteria. A substantive analysis provides a more demanding definition of rationality, which excludes some preferences that would pass a test of coherence. The core of a substantive analysis is an independent assessment of the quality of decision outcomes.

The line between logical and substantive analyses is often fuzzy. For example, the “more is better” rule of dominance is a substantive rule that has the force of a logical principle. A substantive judgment is also implicitly invoked in experimental studies of invariance, where decision makers express conflicting preferences in choice problems that are said to be “the same,” “extensionally equivalent,” or “not different in any consequential respect.” In the mugs experiment, for example, it appears unreasonable for owners and choosers to set very different prices for the same object, because the long-term consumption they will derive from it is presumably the same, and because long-term considerations carry more weight than the transient affect associated with giving up an object. A criterion of utility experienced over time is implicit in this argument.

The research reviewed in earlier sections was explicit in evaluating decisions by a criterion of experienced utility. Various proxies were used to measure this subjective variable. For example, choices made near the moment of consumption were the criterion in evaluating earlier commitments (Simonson [1990]). In other studies, the adequacy of retrospective evaluations and of the decisions...
they support was assessed by applying normative rules (e.g., temporal monotonicity) to real-time records of hedonic experience.

The correspondence of experienced utility and decision utility is commonly taken for granted in treatments of choice. Contrary to this optimistic assumption, two obstacles to the maximization of experienced utility have been identified here. First, preliminary findings suggest that people lack skill in the task of predicting how their tastes might change. The evidence for this conclusion is still sketchy, but its significance is clear: it is difficult to describe as rational agents who are prone to large errors in predicting what they will want or enjoy next week. Another obstacle to maximization is a tendency to use the affect associated with particular moments as a proxy for the utility of extended outcomes. This peculiarity in the cognitive treatment of time explains the importance that people attach to the emotions of transactions, and may cause other forms of myopia in decision making. The use of moments as proxies entails a neglect of duration in the evaluation of past episodes, which has been confirmed in several studies. These results illustrate one particular form of distortion in evaluative memory; there may be others. Observations of memory biases are significant because the evaluation of the past determines what is learned from it. Errors in the lessons drawn from experience will inevitably be reflected in deficient choices for the future.

The rules that govern experienced utility emerge as an important subject for empirical study. For example, research could address the question of how to maximize experienced utility under a budget constraint. Scitovsky [1976] offered an insightful analysis of this problem in his *Joyless Economy*, where he took the position that the maximization of pleasure is a difficult task, which is performed with greater success in some cultures than in others. The process of hedonic adaptation played a central role in his treatment of “comforts,” which suggests that it is pointless to invest resources in objects that quickly lose their ability to give pleasure. Expenditure should be directed to goods and activities that provide recurrent pleasures when appropriately spaced over time. In this light, money may be better spent on flowers, feasts and vacations than on improved durables. A systematic empirical study of the issues that Scitovsky raised is both possible and necessary.

A deeper understanding of the dynamics of the hedonic response is needed to evaluate the welfare consequences of institutions. For example, the course of income changes over a standard academic career appears designed for a pleasure machine that responds well to gradual increments and treats any losses as highly aversive (Frank [1992]; Loewenstein and Sicherman [1991]; Kahneman and Varey [1991]). Another institution that probably delivers improving outcomes over time is the penal system: the well-being of prison inmates is likely to improve in the course of their sentence, as they gain seniority and survival skills. This arrangement is humane, but perhaps less than efficient in terms of individual deterrence. Suppose, for the sake of a provocative example, that prisoners apply a Peak & End rule in retrospective evaluations of their
prison experience. The result would be a global evaluation that becomes steadily less aversive with time in prison, implying a negative correlation between sentence length and the deterrence of individual recidivism. This is surely not a socially desirable outcome. Should shorter periods of incarceration under conditions of increasing discomfort be considered? As this speculative example illustrates, detailed consideration of experienced utility can yield quite unexpected conclusions.

The hedonic criterion of experienced utility is appropriate for some decisions, but it is neither universal nor exhaustive. Rational people may have other objectives than the maximization of pleasure. As Sen has noted, the rationality of decisions is best assessed in the light of "what the person is aiming to do or trying to achieve." At least in principle, a substantive evaluation of individual decisions can be extended to other criterial objectives, such as the achievement of increased personal capabilities, or of a good reputation. As the example of experienced utility illustrates, the investigation of any proposed criterion for decision making must involve three elements: (i) a normative analysis; (ii) development of measurement tools for the evaluation of outcomes; and (iii) an analysis of ways in which decisions commonly fail, by this criterion. Experienced utility is an obvious subject for such a program, but it need not be the only one.

From the point of view of a psychologist, the notion of rationality that is routinely invoked in economic discourse is surprisingly permissive in some respects, surprisingly powerful in others. For example, economic rationality does not rule out extreme risk aversion in small gambles or radical discounting of the near term future, although these attitudes almost necessarily yield inferior aggregate outcomes. On the other hand, rationality is often taken to be synonymous with flawless intelligence. Thus, a critic of the rationality assumption faces the following well-fortified position: (i) a definition of rationality which appears to be overly permissive in some important respects; (ii) a willingness of choice theorists to make the theory even more permissive, as needed to accommodate apparent violations of its requirements; (iii) a methodological position that treats rationality as a maintained hypothesis, making it very difficult to disprove; (iv) an apparent readiness to assume that behavior that has not been proved to be irrational is highly intelligent.

In contrast to the many recent attempts to relax the definition of rational choice, the argument of this essay has been that the definition should be made more restrictive, by adding substantive considerations to the logical standard of coherence. There is compelling evidence that the maintenance of coherent beliefs and preferences is too demanding a task for limited minds (Simon [1955]; Tversky and Kahneman [1986]). Maximizing the experienced utility of a stream of future outcomes can only be harder. The time has perhaps come to set aside the overly general question of whether or not people are rational, allowing research attention to be focused on more specific and more promising issues. What are the conditions under which the assumption of rationality can
be retained as a useful approximation? Where the assumption of rationality must be given up, what are the most important ways in which people fail to maximize their outcomes?

References


Professor Daniel Kahneman
Woodrow Wilson School
of Public and International Affairs
Princeton University
Princeton, NJ 08544
U.S.A.