1 Objective Happiness

Daniel Kahneman

An assessment of a person's objective happiness over a period of time can be derived from a dense record of the quality of experience at each point-instant utility). Logical analysis suggests that episodes should be evaluated by the temporal integral of instant utility. Objective happiness is defined by the average of utility over a period of time. The concept of instant utility must be rich enough to support its role in the assessment of happiness. A purely hedonic concept will not be adequate. The brain constructs a running affective commentary, which evaluates the current state on a Good/Bad (GB) dimension. The commentary has physiological and behavioral manifestations. Although "Good" and "Bad" appear to be mediated by separate systems that can be active concurrently, the description of each moment by a single GB value remains useful. The GB dimension has a natural zero point, "neither pleasant nor unpleasant," which retains its hedonic significance across contexts and permits a measurement of the relative frequencies and durations of positive and negative affect. Comparisons to expectations are an important source of pleasure and pain, but routine experiences are not necessarily affectively neutral. Adaptation to new circumstances has been attributed to a "hedonic treadmill," which reduces the hedonic effect of changes. Some of the evidence for a hedonic treadmill may be due to a satisfaction treadmill, in which the standards that people apply to declare themselves satisfied change. People often assess the well-being effects of states by using the affective value of transitions to these states. Such judgments ignore adaptation. Attempts to estimate the effect of changed circumstances on well-being are susceptible to a focusing illusion in which the weight of the new circumstance is exaggerated. Inferences from preferences to actual hedonic experience are risky. The imbalance of responses to losses and to gains is perhaps more pronounced in decisions than in experience. Retrospective evaluations of episodes give special weight to Peak Affect and End Affect and are insensitive to the duration of episodes. These characteristics of evaluation can yield absurd preferences. Questions about satisfaction with life domains or general happiness are answered by applying heuristics, which are associated with particular biases.

How Happy Was Helen in March? A question is raised in a conversation between two psychologists about a common friend: "How happy was Helen in March?" In the context of an informal conversation, this question would usually be understood and answered with little difficulty. If we know Helen well and saw her often in March, we probably believe we know whether she was happy then, we almost certainly believe that *she* knows whether she was happy then, and with even greater certainty we believe that she knew it then. We also expect our answer to be understood more or less as intended. But we retain this confidence only so long as we remain in the role of intuitive judges in an informal conversation. As soon as we take on the scientific role, we are no longer sure of what the question means, or of the kind of information that we need in order to answer it. The aim of this chapter is to narrow the gap between lay knowledge and professional ignorance. I explore a concept of *objective happiness,* which is an attempt to specify what an objective observer would need to know in order to determine how happy Helen was in March, and the rules for using that knowledge.

A Bottom-up Approach to the Analysis of Well-being

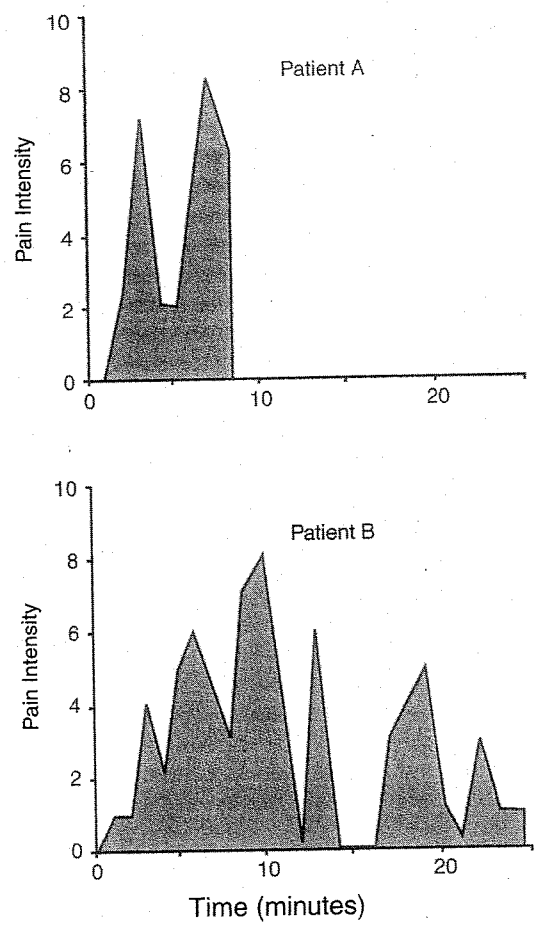
The utterances "I am enjoying this experience and would like it to continue," "Last evening was fun," "I am satisfied with my job," and "I am very happy" all refer to a favorable state of being. All imply a positive value on a broad dimension that will here be labeled GB (for Good/Bad) to avoid the overly intellectual connotation of the commonly used term "evaluation." The four variants of the GB dimension differ in the level of integration to which they refer.

1. Being pleased or distressed is an attribute of experience at a particular moment. I will label this attribute *instant utility,* borrowing the term "utility" from Bentham (1789/1948).[1] Instant utility is best understood as the strength of the disposition to continue or to interrupt the current experience.
2. *Remembered utility* is the global evaluation that is retrospectively assigned to a particular past episode or to a situation in which similar experiences recur. This global evaluation can be expressed in words such as "liked it" and "hated it" or in emotional responses of fear or eager anticipation when a recurrence is likely.
3. Satisfaction questions refer to more inclusive domains of life, such as family life or work.
4. At the highest level of integration we find dimensions such as happiness, or well-being, which encompass all domains of life.

The goal of this chapter is to advance our understanding of the higher level of integration. We want to understand and to be able to assess Helen's happiness. The perspective of the present chapter is bottom-up. It takes the instant utility of the moment as the basic unit of analysis and seeks an objective and normatively justified definition of "true" well-being that is based mainly on information about instant utility.[2] An assessment of Helen's objective happiness in March should be made on the basis of the relevant aspects of her life during that month by applying definite rules to summarize this information in a single value. Helen's own judgment of how happy she was in March is viewed as a fallible estimate of her objective well-being. This conception does not deny the significance of Helen's evaluation of her life. Her thoughts about whether she is currently happy or depressed are themselves causes of pleasure and pain and can be significant if they are frequent and emotionally arousing. In the present framework, however, what Helen thinks about her happiness matters to her "true" or objective well-being only to the extent that her thoughts affect the pleasantness or unpleasantness of particular moments in her life.

Figure 1.1 illustrates the basic approach in an elementary case. The figure presents records of the pain reported by two patients undergoing colonoscopy (Redelmeier and Kahneman 1996). The patients were prompted every sixty seconds to report the intensity of their current pain. They were to use a scale where 10 was "intolerable pain" and 0 was "no pain at all." Later the patients evaluated the experience as a whole, compared it to other unpleasant experiences, and made a hypothetical

FIGURE 1.1 Pain Intensity Reported by Two Colonoscopy Patients



Source: Redelmeier and Kahneman 1996, 4. Reprinted with permission from the International Association for the Study of Pain.

choice between a repeat colonoscopy and a barium enema. We wish to use these data to answer questions such as: "How bad was Patient A's overall experience of colonoscopy?" "Who had the worse experience, Patient A or Patient B?" The approach to these questions (developed in the next section) accepts the patients' ratings of instantaneous pain as valid but does not take their global assessments of the experience at face value. As will be shown later in the chapter, the retrospective evaluations of patients are suspect because they are liable to biases of memory and to a process of evaluation that sometimes violates elementary logical rules. Instead of relying on the patients' judg-

ments, we identify prescriptive principles that should govern the evaluation of episodes. These principles are then applied to evaluate the profiles of particular episodes, such as those illustrated in figure 1.1. The individual's own retrospective evaluation of the experience (its *remembered utility*) is viewed as a fallible estimate of this constructed evaluation, which is called the *total utility* of the episode (Kahneman, Wakker, and Sarin 1997).

This chapter extends the idea of bottom-up construction of global evaluations to levels of integration higher than brief episodes, including judgments of satisfaction with life domains and overall happiness. We distinguish two notions of happiness, or well-being (the two terms are used interchangeably in this chapter). *Subjective happiness* is assessed by asking respondents to state how happy they are. *Objective happiness* is derived from a record of instant utility over the relevant period. The relation between subjective and objective happiness is precisely analogous to the relation between the remembered utility and the total utility of episodes. Like the total utility of a colonoscopy, Helen's objective happiness in the month of March is to be derived according to appropriate rules from a record of her instant utility during that month.

Objective happiness, of course, is ultimately based on subjective data: the Good/Bad experiences of moments of life. It is labeled objective because the aggregation of instant utility is governed by a logical rule and could in principle be done by an observer with access to the temporal profile of instant utility (Kahneman, Wakker, and Sarin 1997). Objective happiness is not to be confused with objective good fortune, which is an assessment of the circumstances of someone's life. All combinations of levels of good or bad fortune, objective happiness or misery, and subjective happiness or misery are possible, and all are probably quite common.

The goals of this chapter are to identify some of the logical and technical problems that need to be solved to turn the measurement of objective happiness into a practical possibility; to identify some of the biases that affect individuals' global judgments of their experiences; and to present an account of the bottom-up approach that is sufficiently clear to advance the discussion of how this approach should be modified, or perhaps to identify fatal flaws that should cause it to be abandoned.

A Logic for Objective Happiness

The conceit of this chapter is that an observer could evaluate Helen's objective happiness in March on the basis of a continuous record of her status on the Good/Bad dimension, along the lines of figure 1.1. This approach is hardly new. More than a century ago, the economist Francis Edgeworth (1881) wrote of using a "hedonimeter" in just this way. A natural way to use such a record is to define the total utility experienced during an interval of time by the temporal integral of instant utility. The temporal integration idea was formulated by Edgeworth (1881), more recent statements of it can be found in Parfit (1984), Broome (1991), and Parducci (1995), and it is invoked implicitly or explicitly in utilitarian analyses (Glover 1990).

The core idea that leads to temporal integration is straightforward, and the colonoscopy example can be used to illustrate it. Obviously, there are two ways of making a painful medical procedure worse: by increasing the level of pain, or by making the procedure longer. Thus, an equivalence can be established between changes of pain intensity and of duration. Furthermore, because duration is measured on a ratio scale in physical units, it is possible in principle to rescale pain intensity in terms of duration.

Kahneman, Wakker, and Sarin (1997) investigated the logical conditions that would justify the temporal integral of instant utility as a measure of the total utility of outcomes. The theory is concerned with profiles of instant utility, such as those illustrated in figure 1.1, that are produced by a subject recording her experiences. Stringent assumptions about the subject's ratings of instant utility are made: these ratings must contain all the relevant information required for its temporal integral to be a plausible measure of the total utility of an extended period. It is also assumed that the scale has a stable and distinctive zero point ("neither good nor bad," "neither approach nor avoid"), and that the measurement of positive and negative deviations from zero is ordinal. The subject's ratings correctly order experiences by the intensity of Good or Bad, but the intervals between ratings may be arbitrary: a pain rating of 7 is reliably worse than a rating of 6, but the interval between 7 and 6 need not be psychologically equivalent to the interval between 3 and 2.

In addition to the subject, the theory involves an observer, who is knowledgeable about the sub-

ject's use of the scale. The observer and the subject may be the same person. The observer's task is to make comparative judgments about utility profiles. These judgments, if they satisfy some rather innocuous axioms,[3] effectively determine the equivalence between the original utility scale and duration. At least in principle, this trade-off can then be used to rescale the reports made by the subject. For example, suppose that the observer judges that one minute of pain at level 7 is as bad as two minutes of pain at level 6. According to the theory, this judgment implies that the original reports of pain should be rescaled, assigning level 7 a value that is twice as high as the value assigned to level 6. If the observer's judgments obey the axioms, the theory asserts that a consistent rescaling is possible, yielding a ratio scale for instant utility that is calibrated by its relation to duration. The rescaling procedure is a close cousin of the method used in medical research to estimate Quality-Adjusted Life Years (QALYs) by establishing equivalences between years of survival in normal health and years of survival at some lower level of health (Weinstein, Fineberg, et al. 1980).

The formal analysis describes a theoretical possibility, not a practical procedure. Its contribution is to clarify the logic that applies to the evaluation of profiles of instant utility. It is important to note that the rule of temporal integration may not apply to the original profiles. It applies only after a rescaling that incorporates a judgment about the equivalence of intensity and duration. Independent of whether or not it is implemented by rescaling instant utility, the principle of temporal integration highlights the importance of duration. The principle is consistent with the intuition that it is imprudent to seek short and intense pleasures that are paid for by prolonged mild distress. The theory also provides an intuitively appealing account of cases in which the consumption of memories matters. It suggests, for example, that the evaluation of the global utility of a safari in Kenya should properly include subsequent episodes of slide-showing and story-telling.

Utility profiles can be used to describe brief episodes, as illustrated by figure 1.1, or longer periods, such as the month of March for Helen. Two families of cases can be distinguished. In some, such as colonoscopies and Caribbean cruises, duration is a relevant characteristic of the outcome that is to be evaluated. It makes sense to say that one cruise is better than another because it is longer. In such cases, as implied by the definition of total utility, the temporal integral of the rescaled profile (the area under the curve) is the appropriate index with which to compare outcomes that vary in duration. In other cases, the duration of the profile is arbitrary: it does not make sense to say that the month of March was better than the first week of March because it was longer. To compare profiles of this type for periods of different length, the appropriate index is the average height of the rescaled profile. Thus, the objective happiness of Helen in March should be measured by the average of the instant utility that she experienced during that period, after appropriate rescaling.

The conception of instant utility is severely constrained by its proposed role in the definitions of total utility and of objective happiness. The concept of what it is that makes a moment good or bad must be sufficiently rich for integration and averaging to be plausible. Philosophical discussions of the measurement of well-being (for a useful introduction, see Brock 1993) remind us of the common intuition that the evaluation of happiness is in part a moral judgment, which invokes a conception of the good life. A physiological indicator that responds strongly to the pleasures of food and sex but shows only a minuscule response to music would be rejected as a measure of instant utility, both on such philosophical grounds and because it would fail to correspond to the intuitions of music lovers about what makes them happy.[4]

What should a concept of instant utility include? The hedonic quality of current sensory experience is the first candidate, of course, but it is not sufficient. The pleasures and pains associated with anticipation of future experience and with remembering the past must surely be counted (Elster and Loewenstein 1992). Other pleasures (and pains) of the mind are to be included as well (Kubovy, this volume). In particular, the measure of instant utility must allow for states of "flow" (Csikszentmihalyi 1990) in which one is so involved in an experience or activity that hedonic value fades into the background of experience. More generally, the index of instant utility should be adequately sensitive to involvement in tasks and activities (Cantor and Sanderson, this volume). And, of course, the notion of a GB value for a moment must be closely related to an assessment of mood, which is interpreted as reflecting the current balance between resources and demands (Morris, this volume). To capture common intuitions about well-being, an adequate measure should also distinguish between activities that have a promotion focus or a prevention focus (Higgins, Grant, and Shah, this volume) and between situations that vary in the extent of personal control (Peterson, this volume).

The diversity of Good and Bad states is intimidating, and the task of constructing a ratio scale measure of instant utility that can be applied to all these states is formidably difficult and perhaps intractable. However, the study of objective happiness can be pursued usefully with much weaker measurements of instant utility. As discussed later, it is not particularly difficult to distinguish good, bad, and neutral moments, and distinguishing a few categories of intensity among good and bad states is probably no harder. And as a first approximation, it makes sense to call Helen "objectively happy" if she spent most of her time in March engaged in activities that she would rather have continued than stopped, little time in situations she wished to escape, and—very important because life is short—not too much time in a neutral state in which she would not care either way. This is the essence of the approach proposed here.

The Good/Bad Dimension

The bottom-up construction of objective happiness from a record of momentary experience requires that each moment be uniquely characterized by a value on the Good/Bad dimension. Two separate assumptions are involved: that the brain continuously constructs an affective or hedonic commentary on the current state of affairs, and that this commentary is adequately summarized by a single value. The first assumption has a fair amount of support; the second is clearly an oversimplification, but perhaps a tolerable one. The two assumptions are discussed in turn in this and in the following section.

The pervasiveness of the GB dimension was noted long ago. Frijda (1986) writes: "According to Wundt (1903), Titchener (1908) and others, 'feelings' are a basic, irreducible kind of mental element. . . . They presuppose the presence of sensations, that is, they presuppose some object the feeling is about. They have the property of subjectivity. They are experienced in one's own subjective response, rather than as asserting a property of the object. They are evaluative. They imply acceptance or nonacceptance of the stimulus or of the experience itself" (179). Frijda cites a participant in an introspective study of pleasant and unpleasant fragrances (Young 1927): "When I say 'pleasant,' it doesn't stand for anything more than I would smell it more if I could." As these quotations from psychological classics illustrate, the GB dimension has two aspects: it involves both an attribute of subjective experience and an action tendency: to go on or to stop.

The idea that evaluation on a Good/Bad dimension occurs continuously and automatically was later developed in several important bodies of research. In their reviews of the literature, Bargh (1997) and Zajonc (1997) both noted the relevance of early studies of the semantic differential (Osgood, Suci, and Tannenbaum 1957), which showed that most stimuli evoke distinctly positive or negative values on a factor of evaluation (marked by the scales "good-bad," "beautiful-ugly," "kind-cruel"). Zajonc (1980, 1997) and Bargh (1997) describe strong experimental evidence for the proposition that every stimulus evokes an affective evaluation, and that this evaluation can occur outside of awareness. As indicated by the famous subtitle of his essay ("Preferences Need No Inferences"), Zajonc (1980) suggested further that evaluation is at least partly independent of the cognitive processing of information about the stimulus. He later argued (Zajonc 1997) that this claim is supported by the discovery of a direct neural pathway that mediates some emotions and bypasses the systems that normally serve conscious processing of information (LeDoux and Armony, this volume).

Lang (1995) has reported studies of the interactive effects of concurrent stimuli, which provide strong support for the existence of a common mechanism that evaluates stimuli as good or bad, pleasant or threatening. Thus, the magnitude of the blink evoked by a burst of loud noise is potentiated in the presence of aversive pictures (for example, poisonous snakes, aimed guns) and apparently inhibited in the presence of pleasing pictures (for example, happy babies, appetizing food). The generality of the evaluative process is also supported by an experiment (Bargh et al. 1996) in which the prior presentation of any positively evaluated word (such as *water*) was found to facilitate selectively the rapid pronunciation of any other positively evaluated word (such as *Friday*).

The close and immediate link between the GB dimension and tendencies to approach or avoid has been demonstrated in several experiments. For example, Bargh (1997) describes a replication of a striking study (Solarz 1960) in which subjects were instructed to eliminate a word from the screen by moving a lever. Half of the subjects pulled the lever toward themselves, half pushed it away. Subjects were relatively faster in pulling a lever toward themselves (approach) in response to an attractive stimulus, and relatively faster in push-

ing the lever away (avoidance) when the word was aversive. Other demonstrations of the links between affective evaluation and movements that express approach or avoidance have been described by Cacioppo, Priester, and Berntson (1993) and by Förster and Strack (1996).

Several physiological correlates of evaluation have been identified (see Ito and Cacioppo, this volume; Davidson 1992, 1994). These include subtle electromyographic changes in facial muscles—with zygomatic activity indicating positive affect and corrugator activity indicating negative affect. In a series of important experiments, Davidson and his colleagues (for a review, see Davidson 1992) have found that differences in the activation of the anterior regions of the left and right cortices are correlated with the quality of experience. Starting in infancy (Davidson and Fox 1989), a predominance of left-sided anterior activation is associated with positive states, whereas a predominance of right-sided activation indicates negative affect. At a still more basic level, there are discussions of specific neural pathways that deal with the computation of overall reward value (see Shizgal, this volume) and specific neurotransmitters that appear to be involved in the control of approach/avoidance tendencies (Hoebel, this volume).

All these lines of evidence, from the introspective to the biochemical, point to the existence of a continuous evaluative process, which manifests itself in physiological responses at several levels, in expressions of affect and in an immediate propensity to approach or to avoid. The continuous Good/Bad commentary is not necessarily conscious. When it is conscious, it is experienced as pleasure or distress, with a corresponding acceptance or rejection of the stimulus. The notions of acceptance and rejection imply that the GB commentary is associated with a disposition to respond both emotionally and instrumentally to an unexpected interruption of an experience: the interruption of a pleasurable activity will elicit frustration and may evoke resistance; the interruption of a painful state will be accepted with relief. The GB commentary has multiple physiological and behavioral manifestations that are potentially available for continuous measurement. The prospects for useful measurement of the momentary GB value are examined in greater detail in subsequent sections.

Is There One GB Value at a Time?

The discussion so far has presupposed that any moment of time can be characterized by a particular value of the GB dimension—positive, neutral, or negative. Doubts would be cast on this assumption by finding that an evaluation can be both Good and Bad at the same time, or by finding that major manifestations of the GB dimension can be dissociated.

The Good and the Bad regions of the GB dimension are subjectively distinguished by different qualities of experience. They also appear to be mediated by different mechanisms (Cacioppo and Berntson 1994). Approach and avoidance are associated with different neurotransmitters (Hoebel, this volume), reward and punishment with distinct neural pathways (Gray 1994), and positive and negative affect with differential lateral activity in the anterior cortex (Davidson 1992, 1994). There is also ample evidence that approach and avoidance tendencies can occur simultaneously or in rapid alternation, generating internal conflict. Furthermore, studies of individual differences suggest that the frequency and intensity of good and bad affect are independent rather than negatively correlated (Diener and Emmons 1985). Cacioppo and Berntson (1994) concluded, from this and other evidence, that evaluation is better described as bivalent than as bipolar (see also Ito and Cacioppo, this volume).

The bivalent nature of the Good/Bad system is not necessarily incompatible with the notion that most moments can be usefully characterized by a single value on a bipolar Good/Bad dimension. A bivalent system yields a bipolar dimension if the separate mechanisms that mediate Good and Bad are mutually inhibitory or reciprocally innervated (Lang 1995) or if the relevant output of the system is the difference between the levels of activity of the two mechanisms (for example, Davidson 1992). Lang's studies of the effects of pleasant pictures on the startle reflex demonstrate the inhibitory connection. Davidson (1992) suggested that the brain may compute both the sum and the difference of the levels of activity in the separate systems that mediate positive and negative affect. He proposed that the GB value corresponds to the difference, and that emotional arousal corresponds to the summed activity in the two systems.

In summary, it appears that most moments of experience can be adequately characterized by a single summary value on the GB dimension. This summary is crude or misleading in some cases: experiences such as those of a straining runner or a spectator watching a tragedy call for more differentiated descriptions. Even in such cases, however, it is usually possible to locate the moment on the Good or on the Bad side of neutral, by applying

the additional criterion of whether an interruption would be welcome or resisted. Would the runner be relieved by an announcement that the race is canceled? Would the spectator welcome the unexpected termination of the performance?

The bipolar nature of the Good/Bad dimension raises some difficult questions. The first such question concerns the relation between pleasure and diminishing pain. Imagine, for example, that you are out in the country during a cold night, inadequately dressed for the torrential rain, your clothes soaked.[5] A stinging cold wind completes your misery. As you wander around you run into a large rock that provides some shelter from the fury of the wind. The event is certainly associated with a reduction of pain. Cabanac (1992) would call the experience of that moment intensely pleasurable, because he believes that the function of pleasure is to indicate the direction of a biologically significant change. However, the experience could also be described as a composite of pleasure and pain, or perhaps as a succession of affective events in which pleasurable relief is quickly followed by a return of (diminished) distress. There is little hope of resolving this problem by introspection, but good reason to believe that the relation between diminishing pain and pleasure will eventually be clarified by studies of relevant brain activity.

Another instructive example, discussed in Kahneman (1992), is the response to a much-wanted salary increase that turns out to be smaller than expected. Casual introspection suggests that, if you expected a raise of $3,000 and received less, any intermediate amount involves both a gain and a loss; there is no intermediate value for which the affective response is neutral. A possible resolution is that this situation resembles familiar examples of bi-stable perceptual organization, such as the Necker cube or Leeper's picture of the wife/mother-in-law. The Necker cube is never seen in two orientations at once, but the orientations dominate in alternation. The perception of a bi-stable figure is best described statistically, by specifying the rate of alternation and the relative proportion of time in which each of the percepts dominates. A similar type of description may prove useful in many situations of affective bivalence and ambivalence.

Observations of hypnotic dissociation reported by Hilgard (1977) suggest that incompatible affective processes can occur in parallel and in relative isolation from each other. Subjects in Hilgard's laboratory were sometimes instructed to maintain a "hidden observer" who would "know what is going on" and who could be reached by tapping the subject's shoulder. In one of the demonstrations, a hypnotized subject was instructed to suck a lemon, which was to be experienced as a delicious orange. As expected, the subject sucked the lemon with every evidence of delight, but when his shoulder was touched he instantly clapped his hand to his mouth, crying, "You have squirted acid in my mouth!" Any simple description of such an experience will be arbitrary to some extent.

Studies of hypnotic analgesia suggest that the hypnotic instruction has its strongest effects on reports of subjective experience and on expressive movements that are under voluntary control. Hilgard states: "The indicators that are essentially *involuntary* (italics in original) have seldom shown consistent reduction under hypnotic analgesia. A subject who is perfectly comfortable and at ease through suggested hypnotic analgesia may still show a rise in heart rate or blood pressure" (Hilgard and Hilgard 1975, 75; see also Berridge, this volume). These observations raise questions about the interpretation of other instances of top-down control of pain (Melzack and Wall 1965). There are well-documented reports of severely wounded soldiers reporting no pain at all during continued fighting and through their evacuation. A compelling film, which often causes observers to feel faint, shows a trepanation operation performed with a blunt-appearing stone on the skull of an awake, impassive member of an African tribe. Do such feats of pain control represent a true reduction of pain or only the inhibition of some of its manifestations? (see Eich et al., this volume; Hilgard 1977).

Toward a Common Metric of GB Value

The references to a single GB dimension in this discussion assume that eating a ripe cherry and watching one's favorite team win the pennant share the important attribute of Goodness, and that arthritic pains and pangs of guilt are both Bad. The project of assessing objective happiness from a record of GB values requires methods of measurement that permit comparisons of GB values across contexts.

An obvious objection to the idea of a common GB metric is that there can be no meaningful comparison of intensity between experiences that differ in their quality. A particular instance of that problem was directly addressed in an important program of research on pain (Melzack 1983; see Eich et al., this volume). The findings confirmed both the existence of qualitative differences be-

tween experiences of pain and the possibility of applying a single scale of global intensity to different kinds of pain. The McGill Pain Questionnaire (MPQ) (Melzack, 1983) consists of twenty separate sets of adjectives that represent different qualities of pain, or different attributes of the experience. The categories are divided into four groups: sensory, affective, evaluative, and miscellaneous. The adjectives are ordered by intensity within each set, and the patient responds to each adjective by indicating whether it applies. For example, scale 1 consists of the labels "Flickering, Quivering, Pulsing, Throbbing, Beating, Pounding." Another sensory scale (7) consists of "Hot, Burning, Scalding, Searing." One of the evaluative scales (16) includes "Annoying, Troublesome, Miserable, Intense, Unbearable." The questionnaire also includes a set of adjectives that describe overall Present Pain Intensity (PPI): "No pain, Mild, Discomforting, Distressing, Horrible, Excruciating."

The exemplary methodology that Melzack and his collaborators applied to develop the MPQ could in principle be extended to other domains. They collected adjectives that are frequently used to describe pain and required judges to group them into sets that could be ordered by intensity, using psychometric methods to select descriptors that represent distinct values on each underlying scale. Finally, they had patients complete the questionnaire and investigated various ways of generating an overall score from the separate scales. A very encouraging result was obtained: a simple sum of ranks across scales yielded a scale that was highly correlated with the PPI measure. This finding sustains the more general hope that it may be possible to measure different kinds of GB experiences on a common scale, without intolerable distortion or loss of information.

Now consider the task of measuring Helen's GB experience during the month of March. Obviously, it is not possible to obtain from her a continuous record of her instant utility. However, the techniques of experience sampling (Stone, Shiffman, and DeVries, this volume) can be used to achieve a useful estimate. Helen might be probed at irregular intervals by a beeper mounted in a special watch, which also displays a scale on which she can select a value that describes the GB value of the moment. To support the measurement of her objective happiness by the average of instant utility, Helen would be required to apply the scale to an extraordinarily broad range of situations and stimuli: embarrassment and a stubbed toe, a gourmet dish and a joke.

Devising a scale that can be applied to many types of stimuli is not difficult. For example, Helen could surely rate qualitatively diverse experiences on a scale anchored, say, on the adjectives "intolerable" and "thrilling."[6] The key question is whether she would be able to use the scale consistently across contexts, thus ensuring that a given rating, say, "Quite Good," corresponds to the same instant utility. To achieve this goal, Helen should be instructed to evaluate her current subjective experience, not the current stimulus. Judging stimuli induces a strong tendency to relate each object to its most natural frame of reference: just as the adjective "tall" does not mean the same thing when applied to a child and to an adult, "quite good" does not have the same hedonic meaning when applied to a breakfast omelette and to a gourmet dish. This tendency can be overcome, however, and there is much evidence of people's ability to adjust their use of scales as required by instructions and by circumstances (Parducci 1995). It is therefore likely that Helen could eventually learn to make GB ratings that reflect the attribute of experience that is shared by good moments of various kinds.

Is there a way to confirm that the GB scale is used appropriately? An economist might wish to anchor the scale in consequential choices, so that GB ratings correspond to preferences. An individual who assigns the same ratings to moments of two different experiences should be equally willing to accept these moments. Unfortunately, single moments are not meaningful units of choice; significant outcomes are normally extended over time and particular moments cannot be segregated. As we shall see later, there are other reasons to reject preferences as the final criterion for the value of experiences.

A more promising approach to the validation of subjective reports of GB values—and more generally, to the measurement of this dimension—may emerge from research on the neuropsychology and psychophysiology of affect (Davidson 1994; Ito and Cacioppo, this volume). It is conceivable, if not likely, that a composite physiological measure of the GB response could eventually be constructed, and that this measurement would be quite highly correlated with subjective experience of pleasure and distress. Continuous records of affective state could possibly be derived from non-invasive measures of localized brain activity, eventually leading to accurate assessments of well-being over time. The movement from science fiction to practical application is likely to be rapid in this domain.

Does the GB Dimension Have a Zero Point?

The question we consider next is whether the zero point of the GB scale ("neither good nor bad") retains the same interpretation as circumstances change in any particular domain, such as food quality, income, or health, and across different domains of experience.

Suppose that Helen's financial circumstances have recently changed for the better, and that her eating habits have changed accordingly. She has graduated from hamburgers and canned tuna to filet mignon and rare tuna steak. Helen's ranking of these items has not changed: even in her days of poverty she preferred tuna steak to tuna salad. However, her talk of food suggests that she now derives less pleasure from food of any given quality: she indicates reduced enjoyment of both superb salad and mediocre steak. In particular, the quality of the food that she labels "neither pleasant nor unpleasant" has also shifted: she now attaches this label to items that she would have called "quite pleasurable" earlier. How are we to interpret what has happened to Helen's hedonic experience of food? Is Helen on a "hedonic treadmill" (Brickman and Campbell 1971; Frederick and Loewenstein, this volume) that completely negates the effect of her improved circumstances?

To interpret Helen's changing hedonic response to food, we compare it to two standard psychological experimental demonstrations: color aftereffects and absolute judgment of length. Imagine a color perception experiment, conducted in two sessions. On each trial of the first session, the subject is initially exposed to strong green light, then to an adjustable mixture of green and red lights. The subject's task is to describe various mixtures on a scale ranging from intense red, through reddish, "neither red nor green" (or white), greenish, and intense green. The second session is similar to the first except for the initial phase of each trial, in which the subject is now exposed to intense red light. A color aftereffect will be observed: the proportion of red light in the mixture that is described as "white" will be much higher in the second session than in the first.

Now consider a two-session experiment in which the subject is exposed to lines of differing length, which are to be described on a scale that ranges from "very long," through "neither long nor short," to "very short." In the first session the subject is exposed to lines ranging from three to twenty millimeters. In the second session, the lines range from ten to fifty millimeters. A context effect will be observed: a line that is described as "neither long nor short" in the first session will be judged "short" or "very short" in the second.

We now consider two questions that are central to the argument of this chapter. Is there an essential difference between the processes that give rise to shifting scales of judgment in the two experiments? If there is a difference, which of the two provides a better model for Helen's changing attitude to food?

Because of the behavioristic tradition of scaling research, the context effects observed in the two experiments are commonly considered similar, if not identical. Both experiments are examples of the oddly labeled absolute judgment task, and standard treatments of this task do not distinguish them (Birnbaum 1982; Helson 1964; Parducci 1995). In fact, the color and length experiments differ in three important aspects.

1. One difference between the two experiments is in the nature of the scales: the scale of length is unipolar, and "short" is roughly equivalent to "not very long"; in contrast, the hue dimension defined by red and green is bipolar, with different sensory qualities on either side of a zero point that is itself distinctive—either "white" or "neither red nor green." The bipolar scale is linked to two distinct mechanisms, which are selectively sensitive to red and green light and linked by an opponent process. A similar structure is found in the temperature sense, a bipolar dimension in which different systems mediate the response to cold and to heat.
2. The similar results observed in the color and length experiments are produced by different processes: color adaptation reflects a change in the sensory mechanism, whereas the context effect observed in size judgments is driven by the requirements of effective communication. Parducci's range-frequency theory (described in the context of an analysis of happiness in Parducci, 1995) explains how respondents adapt their use of the set of labels to the distribution of observations in a way that tends to optimize the informativeness of stated judgments. This is also what people do in adapting labels to categories, as in the familiar example of a large mouse climbing up the trunk of a small elephant.
3. A compelling difference between the experiments is phenomenologically obvious but surprisingly neglected in the scaling literature: subjects in the color experiment will report that the light mixtures that they called

> "white" in the two sessions actually looked alike, whereas the lines they called "neither long nor short" looked quite different. In the length experiment, subjects learn to attach new labels to the unchanging experience of any given stimulus, but in the color experiment it is the experience that changes.[7]

A corollary of these differences is that the neutral value of the scale is appropriately labeled a perceptual zero point in the case of color, but not in the case of length. We are now ready for the question of whether the Good/Bad dimension is more like color or more like length. The answer is unequivocal: in all the aspects listed, the hedonic dimension resembles color more than it resembles size. Like the red-green scale, the hedonic scale is bipolar: pain or distress differs qualitatively from pleasure. There is also evidence for distinct mechanisms that mediate positive and negative affect (see earlier discussion in "Is There One GB Value at a Time"; Cacioppo and Berntson 1994). Finally, hedonic adaptation can be observed without requiring the subject to communicate anything, by observing behaviors of approach and avoidance. Returning to Helen's changing food consumption, the example of the color aftereffect suggests that the hedonic experience of foods that Helen labels (or otherwise treats as) "neither pleasing nor aversive" remains the same as her consumption changes. As Helen adapts to her improving circumstances, however, improving food quality is required to produce this constant experience. This type of adaptation is properly labeled a "hedonic treadmill" (Brickman and Campbell 1971).

Another indication that the zero point of the GB scale is neither arbitrary nor labile is that some experiences never change their sign. Parducci (1995) has noted that there is no context in which cutting oneself shaving will be a pleasant experience. Of course, an individual who normally cuts himself three times every morning might be pleased to observe at the end of a shave that he has cut himself only once—but this is a context-dependent evaluation of the shave as a whole, not an immediate response to the nick of the blade.

The preceding section discussed the ambitious goal of achieving a quantitative measure of the GB dimension that would allow comparisons across contexts. The arguments reviewed in the present section suggest that useful measurements of objective happiness could still be obtained even if precise quantitative measurement of GB values remains elusive. We have concluded that most experiences can be classified as Good, Bad, or neutral with little difficulty and that the neutral affective experience retains the same meaning even when the stimuli that produce it change. It also appears reasonable to assume that the neutral point of the scale is interpersonally comparable, because approach and avoidance have the same meaning for different people (Kahneman and Varey 1991). The intrapersonal stability and interpersonal comparability of the neutral affective experience guarantee the feasibility of at least a crude measurement of the GB dimension. As noted earlier, the relative amounts of time that Helen spent on either side of the GB zero point, or at zero, provide important information about her objective happiness (for similar views, see Diener, Sandvik, and Pavot [1991] and Parducci [1995]).

Norms and Standards

The preceding section distinguished between comparison processes in judgment and in perception. The judgment that a line is long or short arises from an explicit comparison of the current line to a relevant set of stimuli previously experienced. Changing the comparison set changes the judgment, but not the perception of the current stimulus. However, perception itself is also inherently comparative. The perception of a particular mixture of wavelengths as reddish or greenish can be seen as arising from an implicit comparison of the current stimulus to a "memory" of prior stimulation. The adaptation level for hue or temperature is such a memory. Changes in the adaptation level for these modalities are accompanied by changes in perception, as indicated by the possibility of matching experiences (of "white" or "warm") across adapting contexts. We concluded that the zero point of the GB scale is perceptual, and that genuine changes of taste are possible, causing the same stimulus to produce different hedonic experiences and different stimuli to produce identical GB values.

Judgmental comparisons are hardly irrelevant to hedonic life. Indeed, such comparisons are often the *cause* of significant pleasure and pain. The following discussion of the affective consequences of comparison draws on norm theory (Kahneman and Miller 1986). The central idea of norm theory is that reality is continuously experienced in a context of relevant counterfactual alternatives, as each stimulus evokes representations of what it could have been and was expected to be. These repre-

sentations provide a norm to which the evoking stimulus is automatically compared. On most occasions, of course, the stimulus matches its norm and is accordingly experienced as normal. A stimulus that differs sharply from its norm is perceived as surprising or novel. An important tenet of norm theory is *emotional amplification*: the emotional response to abnormal events is enhanced, relative to the response to the same events when they are normal and expected.

Emotional amplification implies that novel events elicit especially strong Good/Bad values. For example, suppose that the newly affluent Helen has switched to a new cereal, which she likes much better than the one she had consumed for a long time. For the first few mornings, the norm that is evoked by the experience of the new cereal mainly consists of memories of the inferior cereal. The positive deviation of present reality from the norm that it evokes surely enhances the pleasure of consuming the new cereal. However, the special pleasure derived from abnormality will inevitably disappear, as the norm gradually changes to reflect Helen's new routine. In the terms introduced in the preceding section, the disappearance of the novelty component of pleasure is a genuine hedonic adaptation. It affects the experience itself, not merely the descriptive labels that would be used to describe it. However, the elimination of novelty pleasure does not imply that the experience of eating cereal is no longer pleasurable. Comparison is not the only source of pleasure, and normality does not imply affective neutrality.

Novelty is only one of several ways in which comparisons induce affect. Comparisons to what might have been can cause counterfactual emotions, such as regret and frustration, guilt and envy (Kahneman and Miller 1986; Roese and Olson 1995). The intense pains and pleasures of comparing oneself to others are central to the lives of many people (Wheeler and Miyake 1992). Comparisons to normative standards and to aspiration levels are particularly important in the context of well-being research, where questions about satisfaction are commonly asked. Satisfaction and the GB dimension are related in two distinct ways. First, high GB values yield satisfaction. Second, the judgment that one is satisfied is an occasion for pleasure as well as a consequence of pleasure. However, the zero point on the satisfaction-dissatisfaction scale ("neither satisfied nor dissatisfied") does not necessarily correspond to a neutral GB value. A gourmet may enjoy a dish that is quite tasty, but not fully satisfactory.

The Hedonic Treadmill

In a landmark essay, Brickman and Campbell (1971) explored the implications of Helson's adaptation level theory for human happiness and for planning the good society. They developed the deeply troubling notion of a "hedonic treadmill": if people adapt to improving circumstances to the point of affective neutrality, the improvements yield no real benefits. Subsequent observations of the apparent lack of effect of increasing real income on satisfaction with income and with other domains of life (Duncan 1975; Easterlin 1974) appeared to provide strong support for the hypothesis of a hedonic treadmill. Perhaps the most dramatic evidence for this hypothesis was the finding by Brickman, Coates, and Janoff-Bulman (1978) that lottery winners were not particularly happy and that paraplegics were much less unhappy than most readers would have anticipated. The distinction between pleasures and comforts that Tibor Scitovsky (1976) developed in his famous book *The Joyless Economy* draws much of its intuitive appeal from the same source. In contrast to pleasures, which are arousing experiences, comforts ultimately produce no significant hedonic experience at all. Related developments include Parducci's psychophysical theory of happiness (Parducci 1968, 1995) and Tversky and Griffin's (1991) endowment/contrast model, in which any pleasant stimulus reduces the pleasure associated with subsequent stimuli of the same kind.

The hedonic treadmill hypothesis and the cluster of findings and ideas surrounding it have been very influential (Frederick and Loewenstein, this volume) because of their links with the familiar themes of the relativity of happiness and the futility of the rat race, and because of the doubts they cast on the welfare consequences of economic progress. However, this seductive idea must be interpreted with caution. For example, a radical version of the treadmill concept might suggest that an experience that is routine and fully expected must become affectively neutral. As noted earlier, however, normality does not imply affective neutrality. Breakfast is almost always pleasant even when thoroughly routinized, and shaving cuts will remain unpleasant even for an inept shaver who cuts himself every morning. The extreme interpretation of the treadmill idea cannot be maintained.

A weaker version of the treadmill hypothesis may suffice to sustain the most provocative implication of Brickman and Campbell's (1971) anal-

ysis, that increases in standard of living do little to improve the human lot, at least above a threshold of adequate per-capita income (see Diener and Suh, this volume). For example, the idea that pleasure is tightly linked to need reduction suggests that different ways of satisfying the same need could eventually yield similar GB values. If this idea is correct, Helen will always derive pleasure from breakfast because she comes to it hungry, but she will ultimately derive no more pleasure from a superior cereal than from an inferior brand.

Another version of the treadmill idea was developed by Headey and Wearing (1992), who argued that the hedonic quality of particular experiences and overall well-being are subject to homeostatic processes that tend to restore a similar distribution of GB values—not necessarily neutral—under varying circumstances. They reviewed evidence that individuals exposed to life-altering events ultimately return to a level of well-being that is characteristic of their personality, sometimes by generating good or bad outcomes that restore this characteristic level. There is other support for the view that happiness—probably both in its objective and subjective varieties—is a personality trait with a large heritable component (Diener and Lucas, this volume). Each individual may be on a personal treadmill that tends to restore well-being to a predetermined set point after each change of circumstances.

The hedonic treadmill hypothesis assumes that a change in objective circumstances causes a predictable change in the GB value of stimuli. Two mechanisms that produce treadmill-like results were illustrated earlier by Helen's response to an improvement in the quality of her food. One suggestion was that adaptation to more palatable food can move the location of the hedonic zero point without altering the palatability ordering, in analogy to color adaptation effects. This mechanism is true hedonic adaptation. The other observation was that changing circumstances yield pleasures and pains that are linked specifically to the contrast between the new circumstances and previous expectations. The amplifying effects of novelty on hedonic experience must eventually disappear as a new routine is established.

A Satisfaction Treadmill

We next consider the possibility of a mechanism that could produce treadmill-like effects without any change in hedonic experience. This mechanism is best labeled a "satisfaction treadmill": it involves a change in the relation between the distribution of GB values and the scales on which individuals report satisfaction and subjective happiness. Consider an individual whose circumstances have changed because of an increase in income or because of a crippling accident. The new circumstances yield a new distribution of Good and Bad experiences in many domains of life. One possible response to this change is to alter the standard by which overall satisfaction with each of these domains is judged: the paraplegic may declare himself satisfied with a leaner diet of pleasure than he required before the accident, and the individual whose income has risen may require a more favorable distribution of GB values to report the same satisfaction as before.

Brickman and Campbell (1971) derived the hedonic treadmill from Helson's (1964) notion of adaptation level. The satisfaction treadmill can be derived from another familiar notion: the aspiration level, which defines a boundary between satisfactory and unsatisfactory achievements. The classic observation about aspiration levels is that they are highly correlated both with real and with expected achievements, though aspirations are generally somewhat higher than expectations (Irwin 1944). It is a commonplace that people's aspiration level for income is moderately higher than their actual income, and research confirms that current income is the single most important determinant of the income that is considered satisfactory (van Praag and Frijters, this volume).

The satisfaction treadmill that is relevant here would operate on the distribution of GB values, not on income as such. The hypothesis is that, as in the case of income, improved circumstances could cause people to require ever more frequent and more intense pleasures to maintain the same level of satisfaction with their hedonic life. In the terms that were introduced earlier, the satisfaction treadmill causes subjective happiness to remain constant even when objective happiness improves.

Of course, hedonic adaptation and changes of aspiration level may both occur. Observed adaptation to new circumstances is the joint effect of the two mechanisms. The relative contributions of the two types of treadmill effects can be studied adequately only by assessing objective happiness. It is remarkable that the necessary research was not done during the first twenty-five years after the formulation of the hedonic treadmill hypothesis—probably because of the absence of a clear distinction between subjective and objective happiness.

As a consequence, the issues raised by Brickman and Campbell (1971) are still unresolved.

The question of whether observed treadmill effects should be attributed to hedonic adaptation or to rising aspirations has important implications for the issues of public policy with which Brickman and Campbell were concerned. The assumption of their analysis was that policies that improve people's circumstances are futile unless they yield an improvement in satisfaction and in subjective happiness. In the present framework, however, it is objective happiness that matters. Policies that improve the frequencies of good experiences and reduce the incidence of bad ones should be pursued even if people do not describe themselves as happier or more satisfied. The recognition that aspiration levels adjust and that people will never be fully satisfied does not mean that they cannot be made (objectively) happier. The implication of this analysis is that the goal of policy should be to increase measures of objective well-being, not measures of satisfaction or subjective happiness.

The Predicted Utility of States: Evaluation by Changes

Duration is a basic and often neglected dimension in the evaluation of the utility of particular outcomes and in the more inclusive assessment of well-being. Pleasure and pain are characteristics of single moments, and instant utility is therefore attached to slices through the stream of experience. However, all hedonically significant outcomes—from the drilling of a tooth to a love affair—are extended over time. Furthermore, the outcomes that are generally considered most significant to well-being are relatively stable states, of wealth, health, employment, or family status. As we will see, however, the task of evaluating such temporally extended outcomes is quite difficult and unnatural.

A thought experiment will place the difficulty of evaluating extended episodes in a broader perspective. Consider an observer in a vision experiment who is exposed on each of a series of trials to an illuminated panel. The trials vary in duration. The luminance of the panel varies both within each trial and across trials. Now consider three tasks that the observer might be assigned:

1. "Assign a value to the current brightness of the panel." This is a standard psychophysical task that requires matching the intensity of a perceptual experience to a value on a scale.
2. "Assign a value to the brightness of the panel on the last trial." This question can also be answered with little difficulty, by visualizing a representative moment of the experience and by assessing the brightness associated with that moment.
3. "Assign a value to the total brightness that you experienced on this trial." This task appears difficult and unnatural, because the total brightness experienced over time is not a quantity for which a perceptual representation is available. Informal observations on tasks of this type suggest that many observers will base their answer primarily on the brightness of the representative moment.

The example highlights two general principles: perception is about the attributes of current events and objects; judgment tasks that require integration of perceptual experience over time are difficult. Both principles apply to evaluations of the utility of temporally extended episodes. Like brightness, the GB value (or instant utility) is an attribute of a moment of experience. Like the total brightness experienced over a trial, the total utility of an episode has no direct perceptual representation and is not easily evaluated. However, there is an important difference between brightness and utility: an inability to evaluate total brightness is probably of little significance to individual adjustment, but an inability to assess the total utility of temporally extended outcomes can be much more important if it causes people to make choices that do not maximize their experienced utility.

The next three sections explore a general hypothesis—labeled *evaluation by moments*—according to which people evaluate the utility of temporally extended outcomes and states by retrieving or constructing a representative moment and evaluating the utility of that moment. The temporal dimension of experience is not directly included in the representations that are evaluated. As a result, the subjective utilities of temporally extended outcomes and states depart systematically from the logic of evaluation, according to which the total utility of an episode is the product of average instant utility and duration (Kahneman, Wakker, and Sarin 1997).

The hypothesis of evaluation by moments must be supplemented by more specific hypotheses concerning the selection or construction of the representative moment. The rules that govern this representation appear to depend on temporal

perspective. When an episode or a state is considered *ex ante,* the initial moment of the episode and the transition to the new state dominate the evaluation. Thus, an evaluation of the subjective utility of changes dominates both the predicted utility and the decision utility of episodes and periods. Other moments are likely to be most salient, however, when the utility of an episode is considered *ex post.* As will be shown, the evidence suggests that the instant utilities of the end of the episode and of its affective peak often dominate its remembered utility.

The hypothesis that people evaluate the utility of future states by evaluating the transitions to these states helps answer a question about the most famous article in the well-being literature. Why did the study of paraplegics and lottery winners by Brickman, Coates, and Janoff-Bulman (1978) become an instant classic and why has it retained its status? The answer is that the main result of this study violates a powerful intuition that paraplegics are utterly miserable, and that this intuition reflects a failure to distinguish appropriately between the state of *being* a paraplegic and the event of *becoming* a paraplegic. The focus on the tragedy of the transition to the paraplegic state is inevitably associated with neglect of the processes of adaptation that were discussed in the preceding section.

To test this hypothesis about the lay theory of well-being, Kahneman and Schkade (1998) asked subjects to evaluate the effect of different features of a new location on well-being. A control group evaluated the features separately (for example, a short and easy commute, or a commute that is long and hard). Several experimental groups were told that a family unexpectedly had to move between two locations that differed in this feature, and they were asked to assess the impact of the new feature on the family's well-being. Separate groups were assigned different temporal perspectives. One group evaluated the impact of the new feature "in anticipation of the move," another "in the first few months after the move," and another group "in the third year." Two additional groups evaluated the impact of the new feature "overall in the first five years," with one of these groups specifically reminded that "when you think about these features, please take a minute to imagine how their influence might change over the years." There were twenty-four pairs of features, which the subjects were instructed to evaluate independently.

The mean ratings of the impact of features on well-being were predicted quite accurately (R = .99) by a simple linear combination of the separate values of the original and of the new feature:

$$V(X \rightarrow Y) = 0.47\ V(Y) - 0.23\ V(X)$$

As expected, this formula reflects the intuition that the contrast with the preceding state affects well-being in a new situation. Quite reasonably, it appears that the direct contribution of the new state is more important than this contrast effect. The equation provides a plausible representation of the response to a change of circumstances.

It should be self-evident that a contrast effect is unlikely to retain its intensity forever. However, the surprising result of the study was that the same formula fit the data for all experimental groups equally well, regardless of temporal perspective. There was no significant indication that the respondents spontaneously realized that the earlier state of affairs would be more important to current well-being in the first few months after a move than three years later. The findings support the hypothesis that people use an estimate of the hedonic impact of a *change* as a proxy for the evaluation of the impact of a new state.

The errors of predicted utility that are produced by the heuristic of evaluating states by moments are generally reinforced by a systematic overweighting of the distinctive aspects of the new state, an effect that has been labeled the *focusing illusion* (Schkade and Kahneman 1998). Gradual shifts of attention are important mechanisms of adaptation to new situations: as time passes, a paraplegic or a bereaved person certainly spends more and more of his or her day attending to aspects of life other than the tragedy. Schkade and Kahneman noted that this aspect of adaptation is particularly difficult to incorporate into predictions of well-being. Consider the questions, "How miserable would you be as a paraplegic?" or, "How happy would you be in California?" It is natural to answer these questions by focusing on the distinctive aspects of life as a paraplegic or in California. However, such a focus inevitably fails to represent the actual experience of people who have had time to adapt and to redirect their attention to other aspects of life. Thus, the impact of any significant new circumstance on well-being is likely to be overestimated when attention is focused on it.[8] Schkade and Kahneman illustrated this by a study that compared the anticipated and the actual effects on well-being of living in California or in the Midwest. The results indicated no difference in the self-reported well-being of students in the two re-

gions (although the Californians were much more satisfied with their climate). However, residents of both the Midwest and California agreed in predicting greater happiness for Californians, probably because they exaggerated the weight of region-specific experiences in everyday life.

The focusing illusion is not a mere artifact in the measurement of well-being; it can have real consequences in people's lives. Some people may be so persuaded that moving to California would increase their well-being that they will actually move there, although it is far from certain that their prediction of the ultimate outcome will prove correct. More generally, an individual may become fixated on the belief that some change will have important consequences for the quality of life, and this belief may then acquire motivating force. Such a fixation of attention could be the origin of many passions, in both the private and the public domains. If the present analysis is correct, of course, most of these passions are built on an illusion.

The Decision Utility of Changes: Gains and Losses

The weight that is assigned to the desirability of an outcome in the context of a decision is called its *decision utility*. Decision utilities are inferred from choices and are used to explain choices. Much of the research on decision-making and on utility has been conducted in a rational and behavioristic tradition, which focuses on observable choices and shuns subjective notions such as experienced utility. It is implicitly assumed in this tradition that the experienced utility of outcomes can be inferred from their decision utility, because rational decision-makers surely know what they will like. In this and the following section, we raise doubts about this inference (see also Berridge, this volume). We also show that the hypothesis of evaluation by moments applies to decision utility, where the relevant moment is the transition from one state to another.

Outcomes can usually be represented and evaluated either as changes or as states. For example, the outcome of a financial transaction can be stated in terms of the amount that was gained or lost (a change of wealth) or in terms of the individual's state of wealth immediately after the transaction. Ever since Bernoulli's classic statement of expected utility theory (1738/1954), standard economic analyses of decision-making have assumed that people evaluate their options by the states of wealth that they could yield. Bernoulli's analysis was a tour de force that anticipated both Weber and Fechner by more than a hundred years. However, his analysis of decision-making was psychologically wrong in a crucial respect: contrary to its main assumption, people do not usually think of outcomes in terms of levels of wealth or income. The analysis that Bernoulli proposed would be valid only if past outcomes did not matter at all—but, of course, they do. Indeed, the utility of a given level of wealth or income depends on the reference to which it is compared: an income of $60,000 does not have the same utility for individuals who recently had incomes of $40,000 or $80,000.

An analysis of risky choice called prospect theory (Kahneman and Tversky 1979, 1984) took a position that is diametrically opposed to the tradition of explaining decisions by attitudes to wealth. In prospect theory, the carriers of decision utility are gains and losses relative to a reference level, which is often the status quo. In this theory, a given state can be assigned quite different utilities depending on the state that preceded it, and quite different states can be assigned approximately the same utility if they represent the same change relative to the reference level.

The centerpiece of prospect theory is the value function for gains and losses of money (illustrated in figure 1.2), and the critical feature of this function is that it is steeper in the domain of losses than in the domain of gains. The differential sensitivity to losses and to gains is called loss aversion. The coefficient of loss aversion is the ratio of the slopes of the value function in the two domains. Figure 1.2 illustrates a value function in which the coefficient is 2.5. The coefficient of loss aversion

FIGURE 1.2 A Typical Value Function

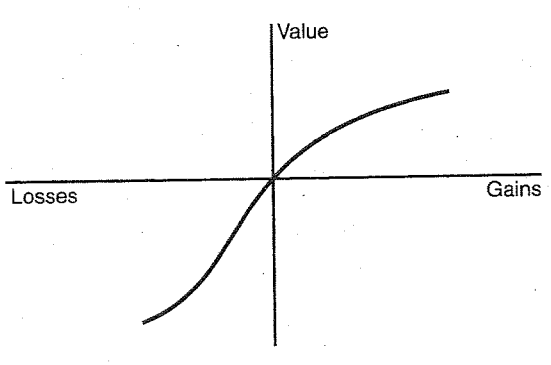
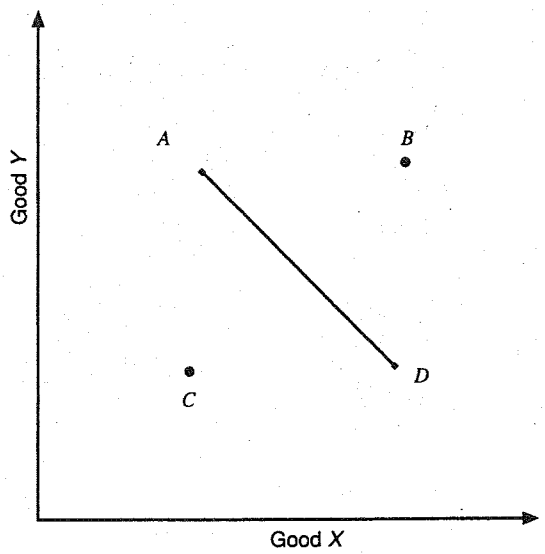


FIGURE 1.3 Multiple Reference Points for the Choice Between A and D



can be estimated, for example, by offering participants a bet on the toss of a coin: they can either lose \$10 or win \$*X*. The factor by which *X* must exceed \$10 provides an approximate measure of loss aversion. The median value in a classroom demonstration is rarely far from \$25.

The coefficient of loss aversion can also be estimated from a comparison of the monetary value that subjects attach to the same object in different situations. In a standard classroom demonstration of this effect, the object is a decorated coffee mug. There are two experimental conditions. (1) Some subjects are given a mug and asked to indicate the minimal amount that would induce them to give it up. (2) Other subjects are given a choice between a mug and an amount of money and asked to indicate the minimal amount of money that they would prefer to a mug. In a typical result, subjects in condition 1 valued the mug at \$7.12, whereas the median subject in condition 2 switched to a preference for money at \$3.50 (Kahneman, Knetsch, and Thaler 1991). This result can be explained by referring to figure 1.3. To read the figure, assume that Good *X* is money and Good *Y* is the mug. The same choice (between points *A* and *D*) is evaluated from point *A* in condition 1, and from point *C* in condition 2. The difference between the states of "having a mug" and "not having a mug" is evaluated as a positive change ("getting a mug") from *C*, and as a negative change ("giving up a mug") from *A*. In both cases, the value of this difference is matched to the value of a monetary gain (Good *X*). Because "giving up" is evaluated on the steep limb of the value function, the gain required to offset it is larger, by a factor of about two.

It is tempting to assume that the properties of decision utility reflect the actual experience of outcomes. Two inferences from figure 1.2 are particularly inviting: (1) the GB value of the reference point is zero; (2) loss aversion reflects a differential sensitivity to pain and to pleasure. Unfortunately, both inferences are dubious.

Prospect theory assigns a value of zero to the reference point that separates positive from negative outcomes. Because the reference point is often the status quo, the theory may appear to incorporate an extreme version of the hedonic treadmill idea, but this inference is unwarranted.[9] The function of figure 1.2 represents the *decision utility* of the gains and losses associated with possible outcomes of the decision at hand; it is silent about the *experienced utility* of the reference situation. Thus, prospect theory is entirely compatible with the conclusion reached earlier (in the discussion of breakfast cereals and shaving cuts) that a state of affairs can be normal without being hedonically neutral.

Another tempting inference from figure 1.2 is that the different slopes of the value function in the positive and in the negative domains reflect differences in the relative sensitivity to pain and to pleasure. However, this inference is also unwarranted. There are two possible interpretations of loss aversion, and they are not mutually incompatible. Loss aversion could represent either a general priority of negative over positive affect in hedonic experience or a deeply ingrained conservative tendency in decision-making. Evidence about the relative importance of the two effects is scarce. However, a psychophysical experiment by Galanter (1992) suggests that the asymmetry between pains and pleasures could well be much smaller in experience than in decisions. Galanter applied the technique of direct magnitude scaling to positive and negative events, including gains and losses of money. He found that the scaled value of gains and losses was a power function of the amount gained or lost. Although there was a difference between the responses to gains and to losses, it was quite small: in a typical experiment the exponents were .54 and .58, respectively, for gains and

for losses. This difference would not account for the extensive loss aversion observed in choice experiments.

The three major conclusions of this discussion of decision utility are that (1) the hypothesis that changes are evaluated as proxies for states holds in the context of decision-making; (2) there is a pronounced asymmetry in the weighting of gains and losses in decisions, but the extent to which loss aversion is also found in experience is not yet known; and (3) more generally, inferences from decision utility to experienced utility should be made with great caution. The next section provides further arguments for the latter conclusion.

Remembered Utility: Episodes and Moments

Consider the following two questions: "How do you feel right now?" and, "How did you feel last night?" Under normal circumstances, we treat answers to these questions with the same easy confidence. Upon reflection, however, the two questions differ greatly in their cognitive demands, and there is a corresponding difference in the authority that should be attached to the answers they evoke. Generating an answer to the retrospective question involves two operations that are not required for the reporting of current feelings: retrieval of a record of GB values from memory, and an act of integrative evaluation that summarizes that record by a single descriptive response.

Earlier discussion ("A Logic for Objective Happiness") proposed temporal integration as the normative principle that should guide the evaluation of a (suitably measured) profile of GB values. However, the hypothesis of evaluation by moments implies that people are unlikely to follow the integration rule, because the temporal integral of momentary sensations is not directly available to conscious awareness. Indeed, several studies have shown that people's intuitive evaluations of their own experiences and of the experiences of others deviate sharply from temporal integration. The participants in these studies generally provided a real-time record of their experience during an episode and later provided a global evaluation of the entire episode, or indicated a choice of which of several episodes they would rather repeat. The situations studied included painful medical procedures, including the colonoscopy study from which figure 1.1 was drawn (Redelmeier and Kahneman 1996); short plotless films of pleasant subjects, such as low-level flying over an African landscape, or of unpleasant subjects, such as an amputation (Fredrickson and Kahneman 1993); immersion of one hand in cold water (Kahneman et al. 1993); and aversive sounds of varying loudness and duration (Schreiber and Kahneman 1998). All these studies share several important features: the participant is essentially passive during a relatively brief episode in which the valence of the experience does not change and the main task is simply to endure the experience and (for some subjects) to evaluate it in real time. The conclusions should not be extended beyond this range of situations.[10]

The retrospective evaluations and choices observed in these studies generally conformed to a simple rule of Peak-End evaluation. Global retrospective evaluations were well predicted by an average of the peak affective response recorded during the episode (in the case of aversive episodes, the worst moment) and of the End value, recorded just before the termination of the episode. For example, in the study from which figure 1.1 is drawn, the correlations between the average of Peak and End ratings and several measures of the patients' global evaluations of their colonoscopy experience ranged from .56 to .67 (Redelmeier and Kahneman 1996). In the study by Fredrickson and Kahneman (1993), subjects were exposed to short films that varied substantially in both duration and affective impact. The mean correlations (within-S) between global evaluations and the Peak-End average were .78 for pleasant films and .69 for unpleasant films. The same rule applies to observers: the physicians who administered the procedures also rated the patient's overall experience according to the Peak-End rule (Redelmeier and Kahneman 1996). Finally, the average of Peak and End discomfort accounted for 93 percent of the systematic variance in an experiment in which subjects rated the total aversiveness of a profile of experience allegedly provided by other people (Varey and Kahneman 1992).

A significant corollary of Peak-End evaluation is duration neglect. Duration was a factor in all the studies mentioned so far, and in others in the same series, but this variable had little or no effect on retrospective global evaluations. In the colonoscopy study, for example, the duration of the procedures varied widely across patients: the mean was twenty-three minutes, and the standard deviation (SD) was thirteen minutes. With Peak and End partialled out, however, duration was not a significant predictor of subsequent global evalua-

tions. Similar results were reported by Fredrickson and Kahneman (1993) in a study of emotionally arousing films. In other studies, the effects of duration were significant but small (Schreiber and Kahneman 1998; Varey and Kahneman 1992, experiment 2). These findings are compatible with the idea of evaluation by moments: people apparently construct and evaluate a representative moment and use the evaluation of this moment as a proxy for the evaluation of the entire episode. Duration is effectively deleted from this representation, in which the Peak and the End are given special weight. Duration neglect is not restricted to human subjects: Mowrer and Solomon (1954) observed that the fear of rats exposed to electric shock depended on the intensity of the shock, not on its duration. Shizgal (this volume) described similar results for electric stimulation of the "pleasure centers" in the rat brain.

The third finding of the studies with human subjects is a robust violation of monotonicity: the retrospective global evaluation of an aversive episode can actually be improved by extending it, provided that the increment yields a lower average of Peak and End values. The participants in an experiment reported by Kahneman and his colleagues (1993) were exposed to two trials of a cold-pressor situation. In the short trial, the subject kept one hand in water at fourteen degrees Celsius for sixty seconds. In the long trial, the immersion lasted a total of ninety seconds. Water temperature was kept at fourteen degrees Celsius for the first sixty seconds, at which point (unbeknownst to the subject) the experimenter caused the temperature of the water to rise gradually from fourteen degrees to fifteen degrees Celsius over the next thirty seconds. Seven minutes after the second trial, the subject was called in for another trial, informed that one of the two previous procedures would be repeated exactly, and given a choice of whether the first or the second trial should be repeated. The robust result of this study is that about 65 percent of participants chose to repeat the long rather than the short trial. The percentage was 80 percent in a subgroup of participants whose real-time ratings indicated a pronounced decline of pain during the last thirty seconds of the long trial.

The results of the cold-water experiment were confirmed and extended in a study using loud aversive sounds as stimuli (Schreiber and Kahneman 1998). Subjects were exposed to pairs of sounds in immediate succession and asked which of the two sounds they would rather hear repeated in a subsequent phase of the experiment. For example, one of the pairs of stimuli that the subjects heard was ten seconds of an unpleasant sound at seventy-eight decibels, then the same sound followed by four additional seconds at a lower intensity (sixty-six decibels). There was a significant preference for repeating the long sounds in such pairs. In a clinical experiment by Katz, Redelmeier, and Kahneman (1996), half of a large group of patients undergoing a colonoscopy were randomly assigned to a condition in which the procedure was briefly extended after the examination was complete, without informing the patient. The colonoscope was left stationary during the added period (about one minute), causing mild discomfort but less pain than many patients had experienced earlier. The prolongation of the colonoscopy, though distinctly unpleasant, yielded a highly significant improvement in subsequent global evaluations of the procedure. A clinical application of such an intervention could be justified if it increased patients' willingness to undergo further colonoscopies when their treatment required it.

The simplest account of this set of results is that subjects form a global evaluation of episodes by the Peak-End rule. As this rule implies, an added period of diminishing pain causes most subjects to retain a more favorable memory of the entire episode. When given a choice, subjects prefer to repeat the trial associated with the less aversive memory. Thus, the remembered utility of past episodes determines the decision utility of repeating them. These results illustrate a general fact of life: except for acts that escape current pain (removing a hand from a flame), the sovereign masters that determine what people will do are not pleasure and pain, but fallible memories of pleasure and pain. The only utility that people (and other organisms) can learn from personal experience to maximize is the utility that they store in memory. Where retrospective evaluations distort actual experience, subsequent preferences are governed by the distorted evaluation, not by the experience.

In the context of measurement of well-being, these results convey a discouraging message: retrospective evaluations of experiences are likely to provide erroneous estimates of the "true" total utility of past experiences. Consequently, Helen's retrospective assessments of the quality of her life cannot be assumed to be valid representations of her objective happiness. However, the analysis also suggests alternatives to methods that rely on global evaluations of the past. One suggestion has already been mentioned on several occasions: because immediate

reports of the quality of experience avoid the difficulties of memory and of integration, experience-sampling methods have significant advantages and should be used whenever possible (Stone et al., this volume).

Experience-sampling studies are cumbersome and expensive, however, and this method will never fully replace retrospective judgments. It will therefore be necessary to develop methods of probing memory that follow the basic principle of human engineering: only ask people to do what they can do well. For example, consider the task of evaluating how happy Helen was yesterday. Helen will probably be able to divide the day into meaningful segments, to estimate the duration of each of these segments with fair accuracy, and to evaluate the average GB value of each segment. The average experienced utility of the entire day is easily derived from these judgments by weighting the average utility of each segment by its duration—an operation that Helen is unlikely to perform well on her own. As this speculative example illustrates, new methods for the assessment of happiness should build on detailed studies of the strengths and weaknesses of evaluative memory.

Heuristics and Biases in Satisfaction and Happiness

People have ready-made answers to many questions about themselves: they know their name, their address, and their party affiliation. But they do not generally know how happy they are, and they must construct an answer to that question whenever it is raised. As they do with other complex questions that must be answered quickly, people are likely to apply simplifying heuristics to the happiness question. As is the case for other simplifying heuristics, the heuristics of satisfaction and subjective happiness are inevitably associated with characteristic biases. There are two general ways of confirming judgmental biases and the heuristics that cause them (Kahneman and Tversky 1996). One method is to compare judgments to true value. In the present context, this method would require comparisons of subjective happiness to independent assessments of objective happiness. An easier way to establish bias is by showing that a judgment is affected by a factor that is normatively irrelevant. Schwarz and Strack (this volume) offer many examples of the latter approach.

Variants of the availability heuristic are natural candidates for answering questions such as, "How satisfied are you with your housing?" or "How happy are you?" or, "How important is this aspect of life to your well-being?" Satisfaction questions can be answered by sampling memories of recent evaluative thoughts or by retrieving and evaluating relevant incidents, which are then compared to a standard or aspiration level. Judgments of satisfaction that are derived in this manner will be overly influenced by recent events and by events that are especially memorable (Tanur 1992). The more general happiness question can also be answered by sampling domains of life and assessing their status. This heuristic yields a bias toward domains that are the focus of current attention, perhaps because satisfaction in these domains is especially high or low.

An earlier section argued that the neutral point of the scale of momentary affect is truly neutral and subjectively distinctive, but this argument does *not* extend to judgments of satisfaction or happiness. The argument was that "neither pleasant nor unpleasant," like "white" or "neither warm nor cold," refers to a distinctive experience at a particular moment. The neutral affective point was said to be determined by a process of hedonic adaptation, which resembles other cases of sensory or perceptual adaptation. In contrast, satisfaction judgments typically refer to a broad domain of life rather than to a single experience, and the anchor of a satisfaction scale is a standard of acceptability or aspiration level, not an adaptation level. The standard for the judgment of happiness is not obvious at all. Most people describe themselves as happy (Diener and Diener 1996), but the meaning of this finding is unclear because the phrase "neither happy nor unhappy" has a distinctly negative connotation.

Other heuristics are at work in judgments of happiness. Schwarz and Strack (this volume) review evidence of large effects of current mood—and even of current weather—on self-reports of subjective happiness. They suggest that the results confirm the role of a "How do I feel" heuristic in individuals' judgments of their happiness (see Schwarz and Clore [1996] for a review). The interpretation is supported by the remarkable finding that the effect of bad weather on reported happiness is reduced or eliminated by drawing the respondent's attention to the weather: this manipulation causes the subject to make a correct attribution of current mood to the weather and reduces the misattribution of current mood to general happiness.

Reports of subjective happiness are highly suscep-

tible to manipulations that attract attention to particular domains of life. In a well-known example, Strack, Martin, and Schwarz (1988) presented the following two questions consecutively in a survey administered to students: "How happy are you?" and, "How many dates did you have last month?" The correlation between these questions varied with the order in which they appeared. The correlation was .12 when the general happiness question came first, but when the dating question came first the correlation rose to .66. Two important conclusions can be drawn from this finding, which has been replicated many times with diverse populations and in a variety of life domains (Schwarz and Strack, this volume). First, people evidently compute an answer to the subjective happiness question on the fly, instead of retrieving a prepared answer from memory. Second, respondents appear to anchor their report of well-being on their satisfaction with any significant life domain to which attention has been drawn. As in the examples of the focusing illusion discussed in an earlier section, selective attention to a significant domain of life produces substantial overweighting of this domain in reports of overall subjective well-being.

In the present framework, the optimal source of data for the assessment of objective well-being would be a continuous record of GB values over time. Although records of affect at the needed level of detail will remain rare for a long time, the adoption of such records as an idealized criterion for the evaluation of well-being has both conceptual and methodological implications. The focus of methodological analyses of self-reported satisfaction and happiness (Schwarz and Strack, this volume) will surely change if these analyses treat reports of subjective happiness as fallible estimates of an objective "true" score. If there is a true score, the operational task becomes the development of practical methods that yield reliable and unbiased measurements of that score. The assumption that objective happiness is well defined and can be measured could be useful in guiding research, even if it is overly optimistic.

Concluding Comment

The present is fleeting, but memories and evaluations of the past endure and populate the mind. When they think about their lives, therefore, people have nothing to work with but memorized assessments and assessments of memories. The central claim of this chapter has been that the scientific study of enjoyment and suffering need not be subject to the same constraint. Real-time measures of experience can be obtained, stored without error, and aggregated to yield a measure of objective well-being that is anchored in the reality of present experience, not in fallible reconstructions and evaluations of the past.

The conception of objective well-being suggests a complex agenda of research, both methodological and substantive. First, there is much to be learned about the various types of real-time measures and how they relate to each other. A second task is to develop methods that minimize the biases of retrospective assessments in order to achieve a measurement of objective happiness that is at once valid and efficient. A combination of methods will eventually be available to characterize the objective well-being of individuals and groups, to determine the true nature of adaptation to new circumstances, to assess enjoyment and suffering in different settings, and to provide a criterion for the evaluation of economic and social policy.

Notes

1. The term "utility" has multiple meanings. Later in the chapter we discuss the concept of *decision utility,* which links utility to choices and preferences rather than to the experience of outcomes.
2. Parducci (1995) offers a similar definition of happiness as "a theoretical summation over separate momentary pleasures and pains" (11).
3. The axioms are: (1) the global utility of a utility profile is not affected by concatenation with a neutral utility profile; (2) increases of instant utility do not decrease the global utility of a utility profile; (3) in a concatenation of two utility profiles, replacing one profile by another with a higher global utility increases the global utility of the concatenation. Peter Wakker has proved the following theorem: "These three axioms hold if and only if there exists a nondecreasing ('value') transformation function of instant utility, assigning value 0 to 0, such that global utility orders utility profiles according to the integral of the value of instant utility over time."
4. Other moral issues relate to the sources of experienced utility. For example, there may be objections to describing Helen as happy if she was maintained in an uncharacteristically euphoric state by mood-altering drugs, if her most significant experiences were delightful hallucinations, or if she lives under such constrained and impoverished circumstances that she has not acquired the normal set of wishes and aspirations (Dasgupta 1993; Sen 1993).
5. The example is Michel Cabanac's.
6. This scale probably incorporates a cultural bias. As several authors have noted (Higgins 1997; Higgins, Grant and Shah, this volume; Russell 1980; Warr, this vol-

ume) Good and Bad states differ on a dimension of arousal. In Wairr's terms, the pleasant states are joy and serenity, the unpleasant states are depression and anxiety. A scale anchored on "thrilling" and "intolerable" implicitly assigns higher value to joy than to serenity.

7. It is also possible to preadapt the two eyes simultaneously to different colors. Preadaptation is followed by a display in which two circles of light are seen next to each other, one of which is shown to the left eye and the other to the right eye. Different mixtures of red and green produce an impression of white in the two patches, depending on the color to which each eye was adapted. No high-level cognitive process is involved.
8. Schwarz and Strack discuss subtle manipulations of mood attribution in which directing attention to an aspect of the current situation that affects mood (for example, the weather) may reduce the effect of the current situation on judged happiness. There is no contradiction between these findings and the focusing illusion, which arises when attention is drawn to significant domains of life.
9. The inference described here as unsound is one I have made in the past. It led me to be insufficiently critical of the hedonic treadmill hypothesis in previous writings (Kahneman and Varey [1991]).
10. The evaluation of episodes of goal-directed activity could well follow a different rule. The affect experienced when the goal is achieved or given up may be much more important to retrospective evaluations than the affect experienced while the activity was going on (Carmon and Kahneman 1996).

References

Bargh, J. A. (1997). The automaticity of everyday life. In R.S. Wyer, Jr. (Ed.), *Advances in social cognition* (vol. 10). Mahwah, N.J.: Erlbaum.

Bargh, J. A., Chaiken, S., Raymond, P., and Hymes, C. (1996). The automatic evaluation effect: Unconditional automatic attitude activation with a pronunciation task. *Journal of Experimental Social Psychology, 32*(1), 104–28.

Bentham, J. (1948). *An introduction to the principle of morals and legislation.* Oxford: Blackwell. (Originally published in 1789)

Bernoulli, D. (1954). Exposition of a new theory on the measurement of risk. *Econometrica, 22,* 23–36. (Originally published in 1738).

Birnbaum M. H. (1982). Controversies in psychological measurement. In B. Wegener (Ed.), *Social attitudes and psychophysical measurement* (pp. 401–85). Hillsdale, N.J.: Erlbaum.

Brickman, P., and Campbell, D. T. (1971). Hedonic relativism and planning the good society. In M. H. Apley (Ed.), *Adaptation-level theory: A symposium* (pp. 287–302). New York: Academic Press.

Brickman, P., Coates, D., and Janoff-Bulman, R. (1978). Lottery winners and accident victims: Is happiness relative? *Journal of Personality and Social Psychology, 37,* 917–27.

Brock, D. (1993). Quality-of-life measures in health care and medical ethics. In M. C. Nussbaum and A. Sen (Eds.), *The quality of life* (pp. 95–132). Oxford: Clarendon.

Broome, J. (1991). *Weighing goods.* Oxford: Blackwell.

Cabanac, M. (1992). Pleasure: The common currency. *Journal of Theoretical Biology, 155,* 173–200.

Cacioppo, J. T., and Berntson, G. G. (1994). Relationships between attitudes and evaluative space: A critical review with emphasis on the separability of positive and negative substrates. *Psychological Bulletin, 115,* 401–23.

Cacioppo, J. T., Priester, J. R., and Berntson, G. G. (1993). Rudimentary determinants of attitudes: II. Arm flexion and extension have differential effects on attitudes. *Journal of Personality and Social Psychology, 65,* 5–17.

Carmon, Z., and Kahneman, D. (1996). The experienced utility of queuing: Experience profiles and retrospective evaluations of simulated queues. Working paper. Durham, N.C.: Fuqua School of Business, Duke University.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience.* New York: Harper and Row.

Dasgupta, P. (1993). *An inquiry into well-being and destitution.* Oxford: Clarendon.

Davidson, R. J. (1992). Anterior cerebral asymmetry and the nature of emotion. *Brain and Cognition, 6,* 245–68.

———. (1994). On emotion, mood, and related affective constructs. In P. Ekman and R. Davidson (Eds.), *The nature of emotion* (pp. 51–55). New York: Oxford University Press.

Davidson, R. J., and Fox, N.A. (1989). Frontal brain asymmetry predicts infant response to maternal separation. *Journal of Abnormal Psychology, 98,* 127–31.

Diener, E., and Diener, C. (1996). Most people are happy. *Psychological Science, 7,* 181–85.

Diener, E., and Emmons, R. A. (1985). The independence of positive and negative affect. *Journal of Personality and Social Psychology, 50,* 1031–38.

Diener, E., Sandvik, E., and Pavot, W. (1991). Happiness is the frequency, not the intensity, of positive versus negative affect. In F. Strack, M. Argyle, and N. Schwarz (Eds.), *Subjective well-being* (pp. 119–40). New York: Pergamon Press.

Duncan, O. (1975). Does money buy satisfaction? *Social Indicators Research, 2,* 267–74.

Easterlin, R. A. (1974). Does economic growth improve the human lot? Some empirical evidence. In P. A. David and M. W. Reder (Eds.), *Nations and Households in Economic Growth* (pp. 89–125). New York: Academic Press.

Edgeworth, F. Y. (1967). Mathematical psychics: An essay on the application of mathematics to the moral sciences. New York: M. Kelly (Originally published in 1881).

Elster, J., and Loewenstein, G. (1992). Utility from memory and anticipation. In J. Elster and G. Loewen-

stein (Eds.), *Choice over time* (pp. 213–24). New York: Russell Sage Foundation.

Förster, J., and Strack, F. (1996). Influence of overt head movements on memory for valenced words: A case of conceptual-motor compatibility. *Journal of Personality and Social Psychology, 71,* 421–30.

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

Frijda, N. (1986). *The emotions.* Cambridge: Cambridge University Press.

Galanter, E. (1992). Utility functions for nonmonetary events. *American Journal of Psychology, 103,* 449–70.

Glover, J. (1990). *Utilitarianism and its critics.* New York: Macmillan.

Gray, J. A. (1994). Three fundamental emotion systems. In P. Ekman and R. Davidson (Eds.), *The nature of emotion* (pp. 243–47). New York: Oxford University Press.

Headey, B., and Wearing, A. (1992). *Understanding happiness: A theory of subjective well-being.* Melbourne: Longman Cheshire.

Helson, H. (1964). *Adaptation level theory: An experimental and systematic approach to behavior.* New York: Harper and Row.

Higgins, E. T. (1997). Beyond pleasure and pain. *American Psychologist, 52,* 1280–1300.

Hilgard, E. (1977). *Divided consciousness: Multiple controls in human thought and action.* New York: Wiley.

Hilgard, E. R., and Hilgard, J. R. (1975). *Hypnosis in the relief of pain.* Los Altos, Calif.: William Kaufmann.

Irwin, F. W. (1944). The realism of expectations. *Psychological Review, 51,* 120–26.

Kahneman, D. (1992). Reference points, anchors, norms, and mixed feelings. *Organizational Behavior and Human Decision Processes, 51,* 296–312.

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,* 401–5.

Kahneman D., Knetsch J. L., and Thaler R. H. (1991). The endowment effect, loss aversion, and status quo bias. *Journal of Economic Perspectives. 5,* 193–206.

Kahneman, D., and Miller, D. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review, 93,* 136–53.

Kahneman, D., and Schkade, D. (1998). Predicting the well-being effect of new circumstances: Changes are proxies for states. Working paper. Princeton, N.J.: Princeton University.

Kahneman, D., and Tversky, A. (1979). Prospect theory: An analysis of decisions under risk. *Econometrica, 47,* 313–27.

———. (1984). Choices, values, and frames. *American Psychologist, 39,* 341–50.

———. (1996). On the reality of cognitive illusions: A reply to Gigerenzer's critique. *Psychological Review, 103,* 582–91.

Kahneman, D., and Varey, C. (1991). Notes on the psychology of utility. In J. Elster and J. E. Roemer (Eds.), *Interpersonal comparisons of well-being* (pp. 127–63). New York: Cambridge University Press.

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

Katz, J., Redelmeier, D. A., and Kahneman, D. (1996). Memories of painful medical procedures (abstract). Fifteenth annual scientific meeting of the American Pain Society, Washington (November 14–17, 1996).

Lang, P. (1995). The emotion probe: Studies of motivation and attention. *American Psychologist, 50,* 372–85.

Melzack, R. (1983). The McGill Pain Questionnaire. In R. Melzack (Ed.), *Pain measurement and assessment.* New York: Raven Press.

Melzack, R., and Wall, P. D. (1965). Pain mechanisms: A new theory. *Science, 150,* 971–79.

Mowrer, O. H., and Solomon, L. N. (1954). Contiguity versus drive-reduction in conditioned fear: The proximity and abruptness of drive reduction. *American Journal of Psychology, 67,* 15–25.

Osgood, C., Suci, G., and Tannenbaum, P. (1957). *The measurement of meaning.* Urbana: University of Illinois Press.

Parducci, A. (1968). The relativism of absolute judgments. *Scientific American, 219,* 84–90.

———. (1995). *Happiness, pleasure, and judgment: The contextual theory and its applications.* Hillsdale, N.J.: Erlbaum.

Parfit, D. (1984). *Reasons and persons.* Oxford: Oxford University Press.

Redelmeier, D., and Kahneman, D. (1996). Patients' memories of painful medical treatments: Real-time and retrospective evaluations of two minimally invasive procedures. *Pain, 116,* 3–8.

Roese, N. J., and Olson, J. M. (Eds.) (1995). *What might have been: The social psychology of counterfactual thinking.* Mahwah, N.J.: Erlbaum.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39,* 1161–78.

Schkade, D. A., and Kahneman, D. (1998). Does living in California make people happy? A focusing illusion in judgments of life satisfaction. *Psychological Science, 9,* 340–46.

Schreiber, C. A., and Kahneman, D. (1998). Beyond the peak-end hypothesis: Exploring the relation between real-time displeasure and retrospective evaluation. Working paper. Princeton University.

Schwarz, N., and Clore, G. L. (1996). Feelings and phenomenal experiences. In T. E. Higgins, and A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 433–65). New York: Guilford.

Scitovsky, T. (1976). *The joyless economy: The psychology of human satisfaction.* New York: Oxford University Press.

Sen, A. K. (1993). Capability and well-being. In M.C. Nussbaum and A. K. Sen (Eds.), *The quality of life* (pp. 30–53). Oxford: Clarendon.

Solarz, A. (1960). Latency of instrumental responses as a function of compatibility with the meaning of eliciting verbal signs. *Journal of Experimental Psychology, 59,* 239–45.

Strack, F., Martin, L., and Schwarz, N. (1988). Priming and communication: Social determinants of information use in judgments of life satisfaction. *European Journal of Social Psychology, 18,* 429–42.

Tanur, J. M. (Ed.). (1992). *Questions about questions: Inquiries into the cognitive bases of surveys.* New York: Russell Sage Foundation.

Tversky, A., and Griffin, D. (1991). Endowment and contrast in judgments of well-being. In F. Strack, M. Argyle, and N. Schwarz (Eds.), *Subjective well-being: An interdisciplinary perspective* (pp. 101–18). Oxford: Pergamon Press.

Varey, C., and Kahneman, D. (1992). Experiences extended across time: Evaluation of moments and episodes. *Journal of Behavioral Decision Making, 5,* 169–86.

Weinstein. M. C., and Fineberg, H. V. (1980). *Clinical decision making.* Philadelphia: W. B. Saunders.

Wheeler, L., and Miyake, K. (1992). Social comparisons in everyday life. *Journal of Personality and Social Psychology, 62,* 760–73.

Young, P. T. (1927). Studies in affective psychology. *American Journal of Psychology, 38,* 157–93.

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist, 35,* 151–75.

———. (1997). Emotions. In D. T. Gilbert, S. T. Fiske, and Lindzey, G. (Eds.), *Handbook of social psychology* (4th ed., pp. 591–632). New York: Oxford University Press.