Evaluation by Moments: Past and Future

Daniel Kahneman
Princeton University

Chapter 38 in:
D. Kahneman and A. Tversky (Eds.)
Choices, Values and Frames
New York: Cambridge University Press
and the Russell Sage Foundation
To appear: summer 2000

September 12, 1999

Evaluation by Moments: Past and Future

Daniel Kahneman[1]
Princeton University

The judgments with which this chapter is concerned are common in daily life. We often have the occasion to evaluate the pleasantness or awfulness of incidents in people's lives; most of us have opinions about what it is like to be old, or physically handicapped, or to live in California; and all of us spontaneously assign grades to events and situations and store evaluations of how good or bad they were, in the form of likes and dislikes. These judgments and feelings have also been produced in the laboratory, with the usual combination of artificiality and improved precision.

The goal of this chapter is to review the evidence for a unifying principle, which accounts for four major conclusions, derived from distinct research paradigms. The first of these conclusions has been supported in many studies of choice.

(i) *The carriers of value* in both risky and riskless choices are gains and losses. The same final state of wealth or endowment is valued differently, depending on its relation to the original state from which it has been reached (Kahneman and Tversky, 1979). In studies of the endowment effect, for example, the outcomes of owning or not owning a particular decorated mug are represented, depending on the current reference point, as *getting* a mug or *giving up* a mug (Kahneman, Knetsch and Thaler, 1991; Thaler, 1980; Tverksy and Kahneman, 1991).

The following sets of conclusions will be documented in the present chapter:

(ii) *Global judgments of fictitious episodes or lives* are very sensitive to trends of improvement or deterioration and are radically insensitive to duration.

(iii) *Retrospective evaluations of affective episodes* are strongly influenced by the affect experienced at singular moments, notably the moment at which affect was most extreme and the final moment. They show little or no sensitivity to duration.

(iv) *Forecasts of the long term effects of circumstances on subjective happiness* tend to neglect the likelihood of adaptation, and therefore to exaggerate long-term benefits and costs of life changes.

All four findings involve the evaluation of an outcome that extends over some time: states that endure for an indefinite term [(i) and (iv)] and bounded episodes, or lives [(ii) and (iii)]. A single psychological process, called *evaluation by moments*, will be invoked here to explain all four findings. Evaluation by moments works as follows: *when an evaluative summary of a temporally extended outcome is required, a representative moment that stands for the entire outcome is selected or constructed; the temporally extended outcome is then assigned the value of its representative moment.* The same general heuristic is applied, in slightly different forms, both to the overall evaluation of past outcomes and to forecasts and decisions about future outcomes. The representative

[1] I thank Dan Ariely, Shane Frederick, Laura Gibson, David Schkade, and Anne Treisman for helpful comments. Dan Ariely and Nathan Novemsky kindly provided data used in Figure 1. I thank Peter Wakker for allowing me to use ideas and sentences that we had fashioned together, Beruria Cohn for allowing me to use her undergraduate thesis data, and Peter McGraw for extremely valuable assistance and useful suggestions.

moment of a past episode is likely to be made up of a collage of impressions and affective reactions associated with salient parts of the experience, most notably its most extreme moment, or peak, and its end. The representative moment that stands for a future state is likely to be the moment of transition to that state. The temporal dimension of outcomes is neglected in both cases. Because moments are slices through time and have no duration, the heuristic of evaluating prolonged outcomes by moments yields evaluations that are insensitive both to the duration of the event and to changes of taste that occur over time.

Section 1 reviews studies of the evaluation of fictitious utility profiles and of the remembered utility of brief pleasant or unpleasant episodes. Section 2 reviews some studies of forecasts about the effect of life circumstances on well-being. Section 3 subsumes the principle of evaluation by moments under a broader judgmental heuristic called judgment by prototype.

1. Evaluating past utility: The Peak-End rule

The research described in this chapter is part of a broader effort to explore *experienced utility* -- a concept that restores Bentham's original definition of utility in terms of hedonic experience. The theory of experienced utility has been developed more fully elsewhere(see chapters 37 and 42). The new terms that will be used in this section are defined below.

Moment-utility: the sign and intensity of affective/hedonic experience at a given moment in time. Moment-utility may be inferred from self-reports or physiological measures.

Utility profile: a verbal or graphic representation of the time course of moment-utility during an episode or over a concatenation of separate episodes.[2]

Evaluation of a utility profile: an observer's judgment about the overall utility of an experience described by a utility profile.

Remembered utility: a subject's own global evaluation of a past episode.

The psychological rules that govern global assessments of episodes have been studied in a diverse set of evaluations: judgments about fictitious profiles of uncomfortable experiences (Varey & Kahneman, 1992) and of entire lives (Diener, Wirtz, & Oishi, 1999); retrospective evaluations of the experience of pleasant or aversive plotless film clips (Fredrickson and Kahneman, 1993), medical procedures (Redelmeier and Kahneman, 1996), painful pressure from a vise (Ariely, 1998), and annoying noises (Ariely and Zauberman, 1999; Schreiber and Kahneman, 2000, expt. 2); measures of the impact of advertisements (Baumgartner, Sujan and Padgett, 1997); and choices about repeating unpleasant experiences (Kahneman et al, 1993; Schreiber and Kahneman, 2000, expt. 3).

[2] An episode is a bounded and compact interval of time defined by its content, e.g., a headache or a vacation (Kahneman, Wakker and Sarin, 1997).

Surprisingly, this body of research shows that the judgment of fictitious utility profiles and the remembered utility of real episodes follow similar rules and violate logic in similar ways. Another surprise is that both types of judgment are well described -- at least in the case of simple affective episodes (e.g., a colonoscopy or a brief film clip of beautiful scenery) -- by a simple *Peak/End* rule: the average of the utility experienced at two singular moments of the episode quite accurately predicts subsequent global assessments.

1.1 Experimental evidence. I first review three experiments in which the temporal profile and the duration of aversive episodes were systematically manipulated. The aim of these experiments was to identify the psychological rules that govern the global evaluation of past (or fictitious**)** episodes:

(i) Participants in an experiment reported by Varey and Kahneman (1992, expt. 2) used a 1-100 scale to evaluate the "total discomfort" experienced by subjects in 48 different conditions of a fictitious experiment. The information for the judgment consisted of a list of ratings of pain on a 0-10 scale, allegedly reported by subjects every 5 minutes during an episode of discomfort. The lists varied in length (from 15 to 35 minutes), in overall level, in trend (ascending or descending) and in the steepness of the trends. Each participant judged 48 profiles, one at a time, after previewing the entire booklet.

(ii) Participants in an experiment described by Schreiber and Kahneman (2000, expt. 2) were exposed to a series of annoying noises, representing different levels of loudness and different temporal patterns. A preliminary study using continuous recordings of annoyance showed that momentary annoyance tracked sound intensity very closely. Profiles of sound intensity were therefore used as a proxy for a continuous measure of experienced utility in subsequent experiments. The "total or overall amount of pleasantness or unpleasantness" of each noise was judged immediately after its termination. The intensity of the noise was varied both within and across trials, in the range of 66db to 80 db, The episodes also varied in duration (from 8 to 24 seconds) and in the shape of the temporal profile.

(iii) Ariely (1998, expt. 2) reported a study in which volunteers were exposed to calibrated pressure from a vise. Several temporal profiles were used (e.g., "Up," in which pressure intensity steadily increased; "Down then Up"; etc.). The duration of the trial was another experimental factor, with three levels: 10, 20, and 40 seconds. Half of the participants provided a continuous record of their pain, which again closely tracked the physical measure of pressure. Subjects experienced considerable pain in this experiment: the average rating of the pain associated with the maximal pressure was about 80 on a scale of 100.

The major empirical findings of these experiments are listed below. As will be seen later, the experimental findings have been confirmed in studies that used correlational designs and choice measures.

Average moment-utility. Other things being equal, the average level of a utility profile accurately predicts retrospective judgments. This is no surprise.

Overall trend. A sequence of increasingly unpleasant experiences is judged much worse than the same experiences in the reverse order. In the Varey/Kahneman study of fictitious profiles, the average rating of a steeply increasing profile of discomfort was 65 on a 100-point scale; the rating of the corresponding profile of diminishing discomfort was 46. The results of Ariely's study of real (and quite severe) pain were strikingly similar: the average rating of a 40-second pattern of steadily increasing painful pressure was 75; the corresponding sequence of diminishing pressure was rated 56.

Duration neglect. The authors of the three experimental studies concluded that the effect of duration was reliable, but small. In the study of pressure pain, for example, a change from 10 to 40 seconds in the duration of the pain increased retrospective ratings by 4.2 scale points, on average -- much less than the 19-point difference between the "Up" and "Down" patterns (Ariely, personal communication, July 1999). Here again, real and hypothetical results matched closely. Varey and Kahneman reported that increasing the duration of fictitious profiles from 15 to 35 seconds raised global ratings by 3.8 scale points -- much less than the difference of 20 points between the "Up" and "Down" profiles.

Violations of dominance: better end. Adding discomfort to an episode cannot truly make it better overall. Nevertheless, adding a period of diminished discomfort to an aversive experience actually improves its global evaluation. Violations of temporal dominance were first observed in judgments of fictitious utility profiles (Varey and Kahneman, 1992). For example, consider the series 2-5-8 and 2-5-8-4, where the numbers refer to reports of pain provided on a 10-point scale every 5 minutes. Although the addition of 5 extra minutes of pain can only increase total discomfort, the mean ratings were 64 for 2-5-8 and 53 for 2-5-8-4. The result was confirmed in an experiment using unpleasant noises (Schreiber and Kahneman, 2000, expt. 2). The retrospective evaluation of a loud noise was improved by adding a period of diminishing annoyance to it: the experience of 16 seconds at 78db, for example, was rated worse overall than the same experience followed by 8 extra seconds at 66db. Because 66db was clearly worse than silence, the longer noise was dominated by the short one but its remembered utility was better.

Violations of dominance: better beginning. Ariely (1998) included in his experiment a "High" condition in which pressure was consistently high (and pain very severe) and an "Up" condition, in which pressure was gradually increased to the same high value. The average ratings for the two conditions were not significantly different. If actual discomfort tracks physical intensity, as his results suggest, this observation also violates dominance.

Additivity of duration effects. The logic of evaluation implies an interaction between duration and intensity: the difference of utility between 30 seconds of pain and 40 seconds of pain should be greater if the pain is intense than if it is mild. However, analyses of duration effects in all three experiments yielded a remarkable and consistent

result: the (slight) effects of duration combined in strictly additive fashion with the effects of other factors.

1.2 The Peak-End rule. Fredrickson and Kahneman (1993) proposed a "snapshot model" of remembered utility -- based primarily on the observation of duration neglect -- which accounts for the first five experimental observations listed above and is readily extended to account for the sixth.

The snapshot model describes how evaluation by moments is applied to the evaluation of past episodes. It asserts that an episode is evaluated by constructing a representative moment -- the snapshot -- which may combine or pool the attributes of separate moments of the actual experience. Fredrickson and Kahneman (1993) proposed that, as a good first approximation, the affective value of the representative moment is a simple average of the most extreme affect experienced during the episode (Peak) and of the affect experienced near its end (End). The affective value of that representative moment, in turn, determines the global evaluation of the entire episode. This simple predictive formula, labeled the ***Peak-End rule***, has proved a good match to the data of several studies. In particular, it accounted for 94%, 86% and 98% of the systematic variance in the three experiments discussed above (respectively, Varey and Kahneman, 1992; Schreiber and Kahneman, 2000; Ariely, 1998).[3]

The Peak-End rule explains the difference between rising and falling trends: in a steadily rising trend the End is as high as the Peak, but in a falling trend the End is lower than the Peak. The Peak-End average is therefore higher in the former case than in the latter. The Peak-End rule also explains the paradoxical finding that adding a period of diminishing discomfort to an aversive episode makes it globally less aversive. Finally, the Peak-End rule entails another paradoxical prediction, which was confirmed in the data of both Varey and Kahneman (1992) and Ariely (1998): similar retrospective evaluations are assigned to an episode in which discomfort increases gradually to a high level, and to an episode in which discomfort is at that level from the beginning.[4]

Two elaborations are required to improve the accuracy of the snapshot analogy. First, because the psychological present has a finite duration, the camera must be imagined as set to a moderately long exposure. No self-report of moment-utility truly describes a single instant: a respondent who is asked about his or her affective state *now* will inevitably respond by evaluating a period of time, which is at least a few seconds

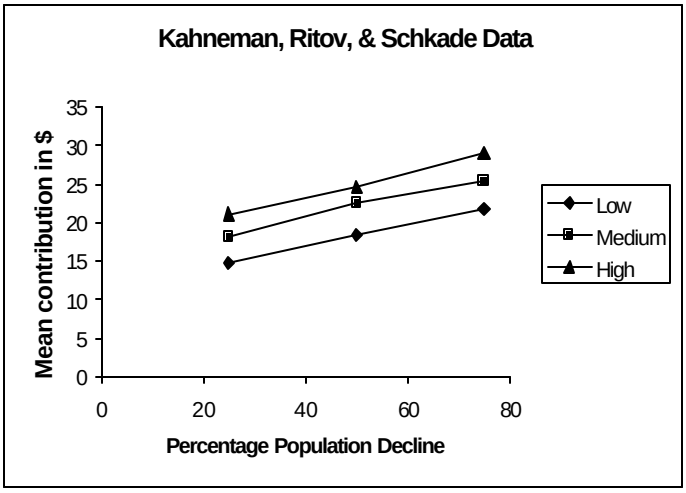
[3] The mean retrospective evaluations are predicted from the Peak and End values of physical sound intensity in the Schreiber/Kahneman study. In Ariely's experiment the predictors are means of subjective judgments; R^2 drops from .98 to .83 when a physical measure is used. The pattern of results suggests that the procedure of using physical measures as a proxy for experienced utility is unsound. However, an experimental condition in which moment-utility is not elicited should also be included, because the elicitation of that measure can be intrusive (Ariely, 1998; Ariely and Zauberman, 1999).

[4] Ariely (1998) inferred from a regression analysis that the final slope is the most important predictor of retrospective evaluation, independent of the effect predicted by the Peak-End rule. This inference is premature, however, because the magnitudes of the different factors cannot be directly compared (Ariely, personal communication, August 1999).

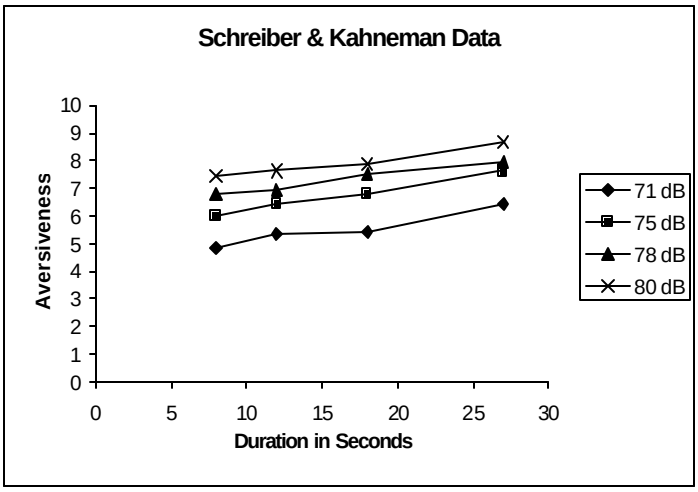
long, sometimes longer. Second, the snapshot should be imagined as including a representation of the affective trend of the experience. As Hsee, Ariely and their colleagues have emphasized, there is an affective response to the trend of affect, which matters to global evaluations (Ariely, 1998; Ariely and Carmon, 1999; Hsee and Abelson, 1991; Hsee, Abelson and Salovey, 1991). In the present conception, hope and fear are incorporated into moment-utility. A measure of moment-utility that is not sensitive to these emotions should be considered seriously flawed.

Figure 1: (A) Willingness to pay to restore damage to species that differ in popularity, as a function of the damage they have suffered (from Kahneman, Ritov, and Schkade, 2000). (B): Global evaluations of aversive sounds of different loudness as a function of duration, for subjects selected for their high sensitivity to duration from Schreiber and Kahneman, 2000). (C): Ratings of probability for predictions that differ in representativeness, as a function of base-rate frequency (from Novemsky and Kronzon, 1999). (D): Global evaluations of episode of painful pressure that differ in temporal profile, as a function of duration (from Ariely, 1998).

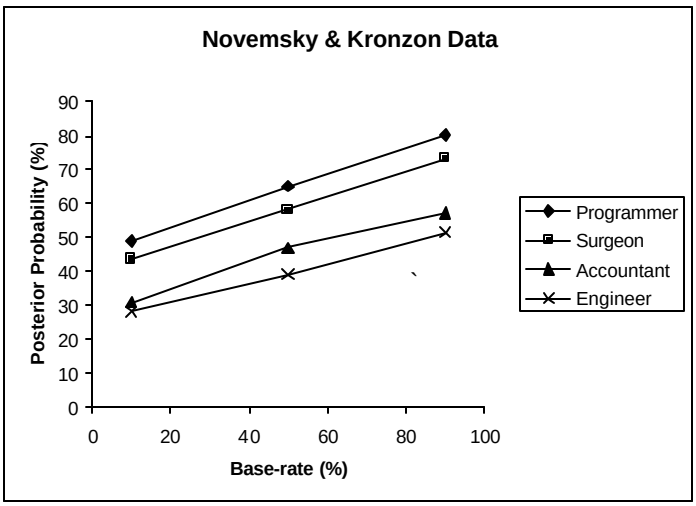
(A)



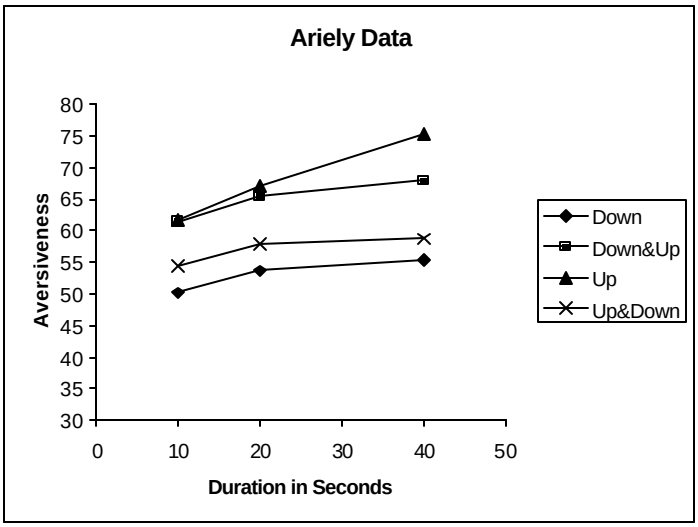
(B)



(C)



(D)



Without further elaboration, the model of evaluation by moments implies complete neglect of duration. As will be seen later this prediction has been confirmed in several studies (Diener, Wirtz, & Oishi, 1999; Fredrickson and Kahneman, 1993; Redelmeier and Kahneman, 1996). However, the principle that it is better for episodes of discomfort to be short rather than long is endorsed by everyone. Duration therefore has some weight in global evaluations when the context both reminds participants of this principle and makes the duration attribute easy to evaluate (Ariely and Loewenstein, in preparation; Varey and Kahneman, 1992, expt. 1). In particular, duration had small but reliable effects in multi-trial experiments which included otherwise identical stimuli (Ariely, 1998; Schreiber and Kahneman, 2000; Varey and Kahneman, 1992). A consistent finding of these experiments was that duration always combined additively with other determinants of global evaluation -- participants appeared to use it as a minor extra feature of each trial, as if they were telling themselves, "This episode is painful and it is also rather long," or, "This episode is painful but it is short." Figure 1 illustrates the additive duration effect observed in two of the experimental studies summarized above. For comparison, it also shows similar results in two other experiments: a study of the

effects of variations of base-rates (Novemsky and Kronzon, 1998) and a study of willingness to pay to restore the population of species that differ in popularity and in the amount of damage they have suffered in a particular habitat (Kahneman, Ritov and Schkade, 1999 [ch. 36]). We return to this remarkably consistent pattern in section 3.

1.3 Additional evidence

Realistic stimuli. The Peak-End rule was initially discovered in an experiment in which participants were exposed to short plotless film clips that varied both in duration and in affective impact (e.g., an amputation, views of a coral reef, etc.). There were two versions of each film, one about three times longer than the other. Each participant saw the long version of eight film clips and the short version of eight others. The mean (computed over Ss) of the correlations (within-S) between retrospective evaluations and the Peak-End average were .78 and .69, respectively for pleasant and unpleasant films. With the Peak-End average statistically controlled, the mean within-S correlation between the remembered utility of film clips and their duration was .06 for pleasant films and -.02 for aversive films -- a striking confirmation of duration neglect. Similar results were obtained with a choice measure.

The Peak-End average also predicted the retrospective evaluation of medical procedures. In Redelmeier and Kahneman (1996) patients undergoing colonoscopy were prompted every 60 seconds to report the intensity of their current pain on a 0-10 scale. Later, the patients provided several measures of remembered utility: they evaluated the total pain that they had experienced during the procedure on a numerical scale, and they also compared the procedure to other unpleasant experiences. The correlation (computed over individuals) between the Peak-End measure and the patient's global evaluation of the procedure was .67. Duration neglect was also observed. The durations of the colonoscopies ranged from 4 to 69 minutes in this study, allowing ample scope for duration to affect remembered utility. Yet the correlation between duration and the patient's subsequent global evaluation was only .03. Physicians' judgments of their patients' experience also showed duration neglect. The physicians were asked, among other questions, whether more anesthetic should have been administered at the beginning of the procedure. The correlation between answers to this question and the duration of the procedure was .05 -- a discouraging finding, perhaps, for potential patients (Redelmeier and Kahneman, 1996).

The colonoscopy study was followed up with a clinical experiment (Katz, Redelmeier and Kahneman, 1997). Half of a group of patients (N = 682) undergoing a colonoscopy were randomly assigned to a condition in which the procedure was extended by about one minute after the examination was complete. (The patients had given advance consent to participation in an experiment, but were not informed of the possibility that the colonoscopy would be prolonged.) The colonoscope was left stationary during the added period, causing mild discomfort, but less pain than many patients had experienced earlier in the operation. As the Peak-End rule predicts, the extension of the procedure, though distinctly unpleasant, yielded a significant improvement in the remembered utility of the procedure. A clinical application of such an

intervention could be justified if it increases patients' willingness to undergo further colonoscopies when their condition requires it.

Evaluations of lives. Diener, Wirtz & Oishi (1999) studied duration neglect and violations of dominance in evaluations of lives. The participants in the experiment were told about a woman named Jen who had, in four basic versions of the story, either an extremely happy or an extremely unhappy life, which ended in a car accident when Jen was either 30 years old or 60 years old. Jen left no descendants. Each participant saw one of the basic versions, and also saw a variation in which Jen lived for five more years before dying in an accident. The five added years were described as less extreme than the preceding years, but their valence was the same: Jen became either less happy or less unhappy than she had been earlier.

Complete duration neglect was observed in answers to both the following questions: "Taking Jen's life as a whole, how desirable do you think it was?" and "How much total happiness or unhappiness would you say that Jen experienced in her life?". Adding thirty years to Jen's life did not change ratings of its desirability, or of the total happiness or unhappiness that she had experienced. As in the Varey/Kahneman study of profiles of discomfort, the overall quality of Jen's life was apparently judged by a weighted average of her happiness over time. Violations of dominance were frequent, especially when the valence of the story was positive: adding five moderately happy years to a very good life made it distinctly worse overall. Lest it be thought that these judgments manifest the folly of youth, a replication in a group of students' parents yielded similar results.

Violations of temporal dominance in choice. The most surprising implication of the Peak-End rule is a violation of temporal dominance: adding an extra period of diminishing discomfort to an unpleasant episode improves its remembered utility, by reducing the aversiveness of the Peak-End average. This hypothesis implies preferences for dominated options in choices that are guided by remembered utility. Such preferences have been observed in two experiments.

Participants in a study reported by Kahneman et al (1993; see Kahneman, 1994 [ch. 42]) were led to expect three trials of a painful experience, and actually experienced two trials, labeled Short and Long. In the Short trial, the subject kept one hand immersed in water at 14°C for 60 seconds. In the Long trial, the immersion lasted a total of 90 seconds. Water temperature was kept at 14°C for the first 60 seconds, at which point (unbeknownst to the subject) the experimenter caused the temperature of the water to rise gradually from 14°C to 15°C over the next 30 seconds. Some time after the second trial, the subject was called in again, 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 several replications of this study was that 65% of participants choose to repeat the Long rather than the Short trial. The proportion rose to 80% when participants who did not indicate a decline of pain during the final 30 seconds of the Long trial were excluded.

Preferences for dominated experiences were also observed in an experiment using loud aversive sounds (Schreiber and Kahneman, 2000, expt. 4). Subjects heard pairs of sounds in immediate succession, were told that one of the sounds would be repeated later, and were offered a choice. In several of the pairs, one of the sounds replicated the other, but with an extra period at a lower intensity (e.g., 10 seconds at 78 db might be followed by 4 seconds at 66db). Here as well, 66% of choices favored the dominated experience. The incoherence of these preferences should be noted: if subjects had been given a button to terminate the sound as soon as they wished, they would certainly have expressed their preference for silence over more of noise. The experiments in which subjects willingly exposed themselves to extra discomfort highlight the tension between two perspectives on experience. The experiencing subject would prefer to stop the unpleasant noise as quickly as possible, but the remembering subject prefers the Long trial to the Short one. Which of these conflicting perspectives should be taken more seriously? This difficult normative question is addressed in more detail in chapters 37 and 42.

1.4 Caveats. The Peak-End rule has only been tested and confirmed in a narrow range of situations. The participants in the studies cited here were always passive during their experience; most of the studies involved aversive experiences; none involved a mix of positive and negative affect. Different rules could apply to other types of episodes. For example, Carmon and Kahneman (1996) suggested that an End rule applies to the evaluation of episodes associated with a goal. On this hypothesis, the affect experienced when the goal is finally achieved or given up will dominate subsequent evaluations of the entire episode. Ariely and Carmon (1999) proposed that prediction of future states is an integral part of overall evaluations. Patients' evaluations in the evening of how much they had suffered during the day appeared to be influenced by the pain they anticipated suffering in the coming night. Ariely and Zauberman (1998) found that the normal preference for improving trends over deteriorating ones is substantially reduced if the same experience is composed of discrete parts. They concluded that the global evaluation of an episode that is broken up into segments is determined by the average evaluation of the separate segments. This important observation calls attention to the crucial role of the parsing of experiences in determining remembered utility: the "final score" of an episode is determined when it is known to have ended, and perhaps only then (Fredrickson, 1991; 1999).

Another caveat applies specifically to studies of choice. Participants in the choice experiments reported in this section apparently let their remembered utility guide their decision. When required to choose between two unpleasant experiences, they selected the one they disliked less. There are, of course, many other strategies of choice, and people do not always make decisions simply by consulting the relative intensities of likes and dislikes. As was noted earlier, decision makers are certainly capable of performing calculations that assign a larger role to duration. Duration is not neglected, of course, in a context that resembles working for pay (Ariely and Loewenstein, in preparation). Informal observations suggest that more of the participants in the cold-water experiment would have selected to repeat the Short trial if they had been given a detailed description of the two trials before making their choice. When a decision is explicitly framed as a choice between a long or a short exposure to pain, people choose correctly. When they go

by what they remember liking, however, their choices are governed by their evaluations of representative moments, which neglect duration.

2. Forecasting happiness: the transition rule

The general hypothesis of evaluation by moments is extended in this section to a new task: the forecasting of the effects of circumstances on well-being. The specific hypothesis that will be examined is that the representative moment which is used to forecast long-term well-being in a new state is the transition to that state. This hypothesis is an immediate generalization of the analysis offered in prospect theory, in which the value attached to possible states of wealth or endowment is determined by the value that is attached to the changes -- gains or losses relative to the status quo -- that may lead to these states. Forecasts of happiness are also made by a ***transition rule***: a prediction of a person's initial reaction to a new situation -- which may be quite accurate in itself -- is incorrectly used as a proxy to forecast the long-term effects of that situation.

The most famous article in the literature on well-being reported that lottery winners are not particularly happy and that paraplegics are not very miserable (Brickman, Coates and Janoff-Bulman, 1978). The study developed the theme of an earlier essay, in which Brickman and Campbell (1971) introduced the notion of a *hedonic treadmill* to describe the general observation that the effects of extreme changes in life circumstances are transient, and ultimately small. The enduring appeal of the empirical study of paraplegics and lottery winners raises a puzzle of its own: If a hedonic treadmill does, in fact, fundamentally govern human well-being, why are we surprised to hear that it does? The answer to this puzzle may lie in the hypothesis that lay forecasts of happiness are made according to the transition rule. Consequently, intuitive predictions about the state of *being* a paraplegic will be dominated by thoughts about the event of *becoming* a paraplegic. Because forecasts that follow the transition rule ignore the evolution of feelings after the initial change, they are often drastically incorrect and generally too extreme. The following two studies illustrate the transition rule in forecasts of well-being.

Table 1 is drawn from a Princeton undergraduate thesis (Cohn, 1999). The respondents (362 adults, recruited by a professional survey firm) were asked to evaluate the well-being of fictitious members of various categories of people, by answering the following question: "Overall, what percentage of the time would you say that Jim is in a good mood, in a bad mood, or in a neutral mood?" The two categories studied in the famous study by Brickman et al. (paraplegics and lottery winners) were included in Cohn's questionnaire. Half of the respondents were asked to assume that the transitional event (becoming paraplegic, winning the lottery) had occurred one month before; the other respondents were told that the event had occurred one year before. Participants also indicated whether they personally knew a paraplegic or a lottery winner, and how well. The results shown in Table 1 are means of the difference between estimates of the percentage of time spent in a good and in a bad mood. Negative values indicate a predominance of bad mood.

Table 1
Mean Mood Difference Scores (%Good mood - % Bad mood)

Category	Paraplegic		Lottery winner	
Knowledge	No	Yes	No	Yes
After one month	-41	-50	58	64
After one year	-37	-19	50	25

The results are clear. Respondents who were not personally acquainted with a lottery winner or a paraplegic were largely insensitive to the time variable. They attributed almost the same level of misery to paraplegics and almost the same level of joy to lottery winners whether a year or a month had passed since the event. The results support the transition rule: In the absence of direct knowledge, people forecast happiness in a long-term state by forecasting (apparently with fair accuracy) the affective impact of the transition to that state. As a consequence, the long-term effects of these life circumstances on well-being are greatly exaggerated. Considerations of adaptation apparently play little or no role in the judgments (see also Loewenstein and Frederick, 1997; Loewenstein and Schkade, 1999; Schkade and Kahneman, 1998).

An experimental study of the lay theory of adaptation was conducted by Kahneman and Schkade (in preparation). Participants were asked to consider a scenario in which a family unexpectedly had to move to a new location. Twenty-four contrasting features of the current and the new location were briefly described. The commute in the new location, for instance, would be short and easy, compared to the current commute, which was long and difficult. The respondents' task was to evaluate the effect of each new feature on the well-being of the family, on a scale that ranged from -5 (Extremely Negative) to 0 (Neutral) to +5 (Extremely Positive). The independent variable was the temporal perspective that the respondents were asked to assume. Participants in one condition of the experiment received the following instructions, which were repeated saliently on each page of the questionnaire: "For each item below, please indicate how this feature of the new location would affect Mr. and Mrs. A's well-being <u>during the first few months after the move</u>, while they are still becoming familiar with the new location."

Diferent groups of participants evaluated the impact of the new features

(i) "in anticipation of the move";

(ii) "in the first few months after the move";

(iii) "in the third year after the move";

(iv) "overall, for the first five years after the move";

(v) "overall, for the first five years after the move," with the following additional reminder: "When you think about these features, please take a minute to imagine how their influence might change over the years."

Table 2
Estimated Impact of a Change on Well-Being

	Before	Third Month	Third year	Five years	Five years*
Cost of living (high to low)	3.90	3.49	3.83	3.94	3.93
Obnoxious relative (near to far)	2.92	2.89	2.64	2.69	3.01
Long winter (long to short)	2.35	2.34	2.29	2.12	2.41
Means (24 items)	2.76	2.71	2.77	2.71	2.84

Table 2 presents mean ratings of changes for the better for three pairs of features (n = 125 per cell), as well as the overall means for all 24 pairs. Precisely the same pattern of results was found in ratings of changes for the worse. The only difference was that ratings of the effect of changed circumstances on well-being were consistently higher for improvements than for deteriorations, by about 1.2 scale points on average.[5]

Here again, the results are straightforward: the manipulation of temporal orientation had no significant effect at all, although the size of the sample provided ample statistical power. The respondents apparently applied a theory of well-being in which the joy of moving away from an obnoxious relative will never pall and the distress of having to live without good produce will never subside. The reminder provided in condition (v) did not change the results. There was no indication that respondents were sensitive to the distinction between the short-term impact of a new circumstance and its ultimate effect on well-being.

A within-S design provides a much more sensitive test of whether people believe in a treadmill effect. We therefore included a condition in which respondents were required to evaluate (on the same line of the questionnaire) the effect of a new circumstance on well-being, both for the first few months after the move and during the third year. The results were meager. Predicted well-being effects decreased substantially (by 0.50 scale points or more) only for 4 of 48 new features: freedom from an obnoxious relative, mediocre produce, humid summers, and long, cold winters. Overall, the within-S experiment showed no evidence of a general belief in a treadmill effect: the average predicted impact of a change (disregarding sign) was exactly the same (within .02 of a scale point) for the first three months and for the third year.

Additional analyses relied on judgments provided by control groups, whose members rated the well-being effect of each of the 48 separate features. The mean ratings of the separate features in each pair were used to predict the well-being effects of a change from one of these features to the other. The effect of a change from feature X to

[5] The prevailing optimistic bias is an instance of the more general effect that Cacioppo and Berntson (1994) have called a positivity offset.

feature Y was predicted with high accuracy (R^2 = .98 for all temporal orientations) by the following formula:

$$V (X \text{ --> } Y) = 2\ V(Y) - V(X),$$

where V (X --> Y) is the predicted effect of the new feature, and V(X) and V(Y) are ratings of the separate features by the control group. The formula reflects the belief that contrast with an earlier state affects well-being in a new state, and it reasonably assigns more weight to the new feature than to the past. Thus the equation provides a plausible representation of the utility of a *change* from one state to another.[6] The noteworthy result of the study is that the very same formula predicted the judgments for all temporal perspectives equally well. This result supports the hypothesis that respondents applied the transition rule to predict well-being in a new situation -- and did so even for a situation that was no longer new. Although we had formulated the study with that hypothesis in mind, the total neglect of adaptation in our subjects' judgments was still surprising. We had expected to find at least a trend indicating an effect of time, but we found none. In the within-subjects condition, we confidently expected substantial contrasts in direct comparison of the first few months with the third year, but the differences we observed were negligible, except for a few pairs of features.

The transition rule helps explain why the Brickman *et al.* study of paraplegics and lottery winners is an enduring classic: its findings are counter-intuitive and perennially surprising, except perhaps to those who have a paraplegic or a lottery winner among their personal acquaintances. The present results are also relevant to a deeper puzzle: Why do people strive so eagerly to improve their circumstances, when a treadmill awaits them? Our findings suggest a simple answer: This essential fact of life is not generally known, perhaps because there are no good opportunities to learn it. Transitions are well-defined events, which often evoke strong emotions and therefore provide the feedback needed to develop the skill of forecasting affective reactions to changes -- one's own and those of others. In contrast, feedback from states is always delayed and its attribution is elusive (Loewenstein and Schkade, 1999). Thus, people can learn to predict reactions to changes without ever learning to forecast how feelings will evolve over time in the absence of change.

The hypothesis of a hedonic treadmill can be applied even when there is no transition from one state to another. In a steady state, categories and groups of people should not differ in happiness, if all have adapted to their circumstances. A study reported by Schkade and Kahneman (1998) deals with one illustrative problem to which this hypothesis applies: Are people happier in California than elsewhere? There is evident tension between the treadmill hypothesis and the widespread belief that California is a highly desirable place to live. Why should living in California be so desirable if it offers no greater happiness? Students at four state universities (two in

[6] A different pattern was observed in a group of respondents who used the same questionnaire to predict the difference in well-being between two families that differ in the relevant feature. Unlike predictions, judgments of well-being differences assigned equal weight to the two features, and were precisely symmetric around zero.

California and two the Midwest; total N = 1070) completed a questionnaire that probed their overall life satisfaction and their satisfaction with separate domains of life. Other participants (total N = 890) completed the same questionnaire with the answers that they attributed to a student at another university, described by the phrase "with values and interests similar to yours." The target person was said to be located at another state university, either in the same region or in the other region.

The results tell a simple story. Students in both regions believed that Californians would report substantially higher life satisfaction, and statistical analysis traced the differential forecasts to attitudes to climate. Participants also expected Californians to be much more satisfied with their climate. This prediction was correct: the Californians liked their climate and the Midwesterners despised theirs -- but climate was not, in fact, an important determinant of overall life satisfaction. With ethnicity controlled, self-reported life satisfaction was identical among students in the two regions. The findings provide a straightforward demonstration of an illusion in the forecasting of happiness. However, an argument that adds much complexity to this simple story is presented in chapter 37.

3. The Principle of Judgment by Prototype

The principle of evaluation by moments asserts that people evaluate the utility of temporally extended outcomes and states by retrieving or constructing a representative moment and by evaluating the utility of that moment. This heuristic of valuation leads to violations of logic, because the temporal dimension of experience is not directly included in the representations that are evaluated. Evaluation by moments provides a compact account of much of the research that has been reviewed here; it also subsumes the earlier conclusion that gains and losses are the main carriers of decision utility in choice.

Evaluation by moments is itself subsumed under a broader principle that has been labeled ***judgment by prototype***. The following propositions (from Kahneman, Ritov and Schkade, 1999 [ch. 36]) introduce this idea:

1) *People hold stored prototypes of many categories. They also form prototypes or representative exemplars for new categories and sets that they encounter.* This proposition refers to the vast separate literatures on the role of exemplars in categorization, on stereotyping, on the role of vivid examples in comprehension of abstract ideas, and on the representativeness heuristic in judgment. In the context of this chapter, the prototype is the representative moment that stands for a temporally extended outcome.
2) *In* ***judgment by prototype,*** *a global judgment of a category or set is determined primarily by the relevant properties of its prototype.* This proposition is central to the idea that a representativeness heuristic is applied in statistical judgment and in intuitive predictions. It is also relevant to situations in which the stereotype that represents a group in people's minds may be a small and statistically unrepresentative minority. Environmental problems, such as the deaths of numerous birds who drown in oil spills or in oil ponds, may be represented by one vivid image, for example a single bird drowning in oil (see ch. 36 for further

detail). The argument of the present chapter is that global evaluations of past episodes were represented by an affectively representative moment -- the Peak-End average; future states were represented by the moment of transition.

3) *When the size of the set is logically relevant to its valuation, judgment by prototype leads to a bias of* ***extension neglect****: Unless attention is specifically directed to it, the size of the set has little or no influence on its valuation.* Extension neglect is illustrated by the neglect of sample-size information in statistical judgment (Tversky and Kahneman, 1971; Tversky and Griffin, 1992), by the neglect of base-rate information in some tasks of intuitive prediction (Kahneman and Tversky, 1973), and by the observation that the willingness to pay for solutions to public problems is insensitive to the quantitative aspect of the problems (Kahneman, Ritov and Schkade, 1999 [ch. 36]). Duration neglect is, of course, a special case of extension neglect.
4) *Extension neglect is neither universal nor absolute. An* ***additive extension effect*** *is observed when the information about extension is both salient and readily interpretable: the effects of the valuation of the prototype and of the size of the relevant set are additive. This pattern violates normative rules that require non-linear combination of the two types of information*. Figure 1 above illustrated the additive duration effect observed by Schreiber and Kahneman (2000), along with two figures drawn from studies in different paradigms: a study of the additive effects of representativeness and base-rate (Novemsky and Kronzon, 1998) and a study of the additivity of the appeal of a species and of the extent of damage to it in determining willingness to pay for restoration efforts (Kahneman, Ritov and Schkade, 1999 [ch. 36]). The same procedure was applied in all three cases to induce sensitivity to extension: the variable was manipulated across trials in a within-subject experiment. The additive duration effect is evidently an instance of a much broader phenomenon, in which people appear to add effects where they should multiply them.[7] As noted earlier, this observation implies that the extension attribute is not incorporated in the basic representation (the prototype) that is evaluated. It is added as a separate constituent of evaluation when specifically primed. Similar results have been observed in many other studies of perception and judgment (Anderson, 1996).

The similarities observed in this highly diverse set of judgment tasks are too consistent to be coincidental. It is reasonable to conclude that judgment by prototype is a general-purpose heuristic, which is applied to diverse cognitive tasks. Whenever possible, it seems, people tend to think in the language of prototypes and exemplars. An important advantage of this heuristic is that it retains only the basic template or schema for a single case, with its specific pattern of slots and default values, and thus undoubtedly achieves considerable economy of processing. Occasionally, however, judgment by prototype leads to mistakes, such as conjunction errors in judgment (Tversky and Kahneman, 1983) or violations of temporal dominance in choice. In the problems that were discussed here, evaluation by moments led to economical but biased

[7] This observation should not be misconstrued as denying people's abilitiy to multiply. In some situations, such as negotiating for pay, people will have no difficulty multiplying an hourly rate by the time they expect to work.

judgments of the affective past, and to systematically incorrect forecasts of the affective future.

References

Anderson, N. H. (1996). *A functional theory of cognition*. Mahwah, NJ: Erlbaum.

Ariely, D. (1998). Combining experiences over time: The effects of duration, intensity changes & on- papers line measurements on retrospective pain evaluations. *Journal of Behavioral Decision Making,* 11, 19-45.

Ariely, D., & Carmon, Z. (1999). Gestalt characteristics of experienced profiles. *Journal of Behavioral Decision Making,* Forthcoming.

Ariely, D., & Loewenstein, G. (1999). Encoding sequences and conveying information in ratings and choice: The case of sequences' characteristics and duration. Working paper.

Ariely, D., & Zauberman, G. (1999). On the making of an experience: The effects of breaking and combining experiences on their overall evaluation. *Journal of Behavioral Decision Making*, Forthcoming.

Baumgartner, H., Sujan, M., & Padgett, D. (1997). Patterns of affective reactions to advertisements. The integration of moment-to-moment responses into overall judgments. *Journal of Marketing Research,* 34, 219-232.

Brickman, P., & Campbell, D.T. (1971). Hedonic relativism and planning the good society. In M.H. Apley (Ed.), *Adaptation-level theory: A symposium* (pp. 287-301). New York: Academic Press.

Brickman, P., Coates, D., & Janoff-Bulman, R. (1978). Lottery winners and accident victims: is happiness relative? *Journal of Personality and Social Psychology,* 37, 917-927.

Cacioppo, J.T., & Berntson, G. (1994). Relationships between attitudes and evaluative space: A critical review with emphasis on the separability of positive and negative substrates. *Psychological Bulletin,* 115, 401-423.

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

Cohn, B. (1999). The lay theory of happiness: Illusions and biases in judging others (Undergraduate dissertation, Princeton University).

Diener, E., Wirtz, D., & Oishi, S. (1999) The James Dean Effect: Peak-End processes in the perception of quality of life. Working paper, University of Illinois at Champaign-Urbana.

Fredrickson, B.L. (1991). Anticipating endings: An explanation for selective social interaction (Doctoral dissertation, Stanford University, 1990). *Dissertation Abstracts.*

Fredrickson, B.L. (1999). Extracting meaning from past affective experiences: The importance of peaks, ends, and specific emotions. *Cognition and Emotion*, Forthcoming.

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

Hsee, C. K. & Abelson, R. P. (1991). Velocity relation: Satisfaction as a function of the first derivative of outcome over time. *Journal of Personality & Social Psychology*, 60, 341-347.

Hsee, C. K., Abelson, R. P., & Salovey, P. (1991). The relative weighting of position and velocity in satisfaction. *Psychological Science*, 2, 263-266.

Kahneman, D. (1994). New challenges to the rationality assumption. *Journal of Institutional and Theoretical Economics*, 150, 18-36.

Kahneman, D., Fredrickson, B.L., Schreiber, C.A., & Redelmeier, D.A. (1993). When more pain is preferred to less: Adding a better end. *Psychological Science,* 4, 401-405.

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

Kahneman, D., Ritov, I., & Schkade, D. (1999). Economic preferences or attitude expressions? An analysis of dollar responses to public issues. *Journal of Risk and Uncertainty,* Forthcoming.

Kahneman, D., & Schkade, D. (1999). Predicting the well-being effect of new circumstances: Changes are proxies for states. Working paper, Princeton University.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80, 237-251.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decisions under risk. *Econometrica*, 47, 313-327.

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, & Kahneman D. (1997). Memories of painful medical procedures. [Abstract]. American Pain Society 15th Annual Scientific Meeting

Loewenstein, G., & Frederick, S. (1997). Predicting reactions to environmental change. In: M. Bazerman, D. Messick, A. Tenbrunsel and K. Wade-Benzoni (Eds.), *Environment, ethics and behavior*. (pp. 52-72). San Francisco: New Lexington Press.

Loewenstein, G., & Schkade, D. (1999). Wouldn't it be nice: Predicting future feelings. In D. Kahneman, E. Diener, & N. Schwarz (Eds.), *Well being: The foundation of hedonic psychology* (pp. 85-108). New York: Russell Sage.

Novemsky, N., & Kronzon, S. (1999). How are base-rates used, when they are used: A comparison of Bayesian and additive models of base-rate use. *Journal of Behavioral Decision Making*, 12, 55-69.

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

Schkade, D. & Kahneman, D. (1998). Does living in California make people happy? A focusing illusion in judgments of life satisfaction. *Psychological Science*, 9, 340-346.

Schreiber, C.A., & Kahneman, D. (2000). Determinants of the remembered utility of aversive sounds. *Journal of Experimental Psychology: General*, Forthcoming.

Thaler, R.H. (1980) Towards a positive theory of consumer choice. *Journal of Economic Behavior and Organization*, 1, 39-60.

Tversky A., & Griffin D. (1991). Endowment and contrast in judgments of well-being. In F. Strack, M. Argyle, & N. Schwarz (Eds.), *Subjective well-being: An interdisciplinary perspective* (pp. 101-118). Oxford: Pergamon Press.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, 76, 105-110.

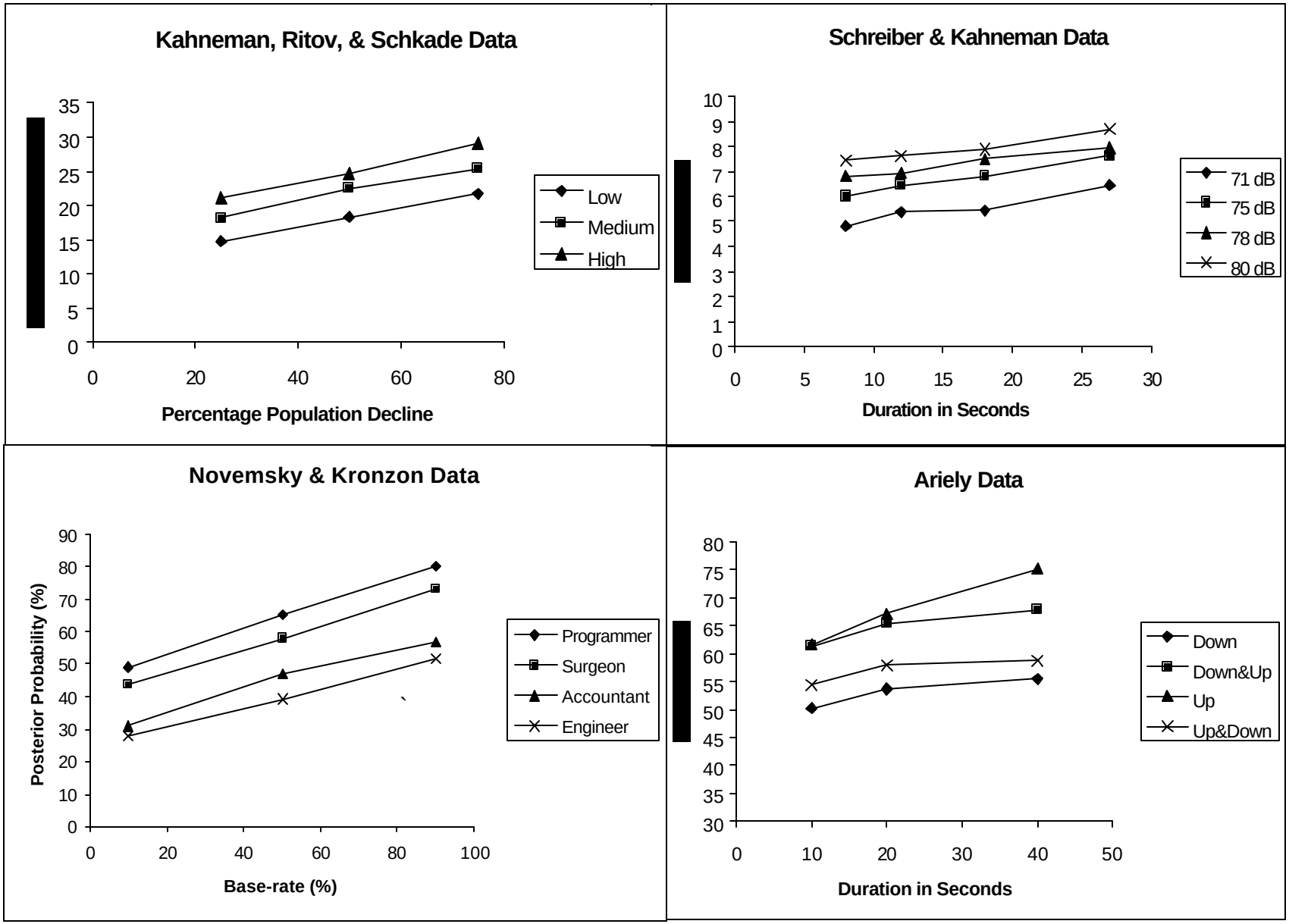
Tversky, A., & Kahneman, D (1983). Extensional vs. intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90, 293-315.

Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. Quarterly Journal of Economics, 106, 1039-1061.

Varey, C., & Kahneman, D. (1992). Experiences extended across time: Evaluation of moments and episodes. *Journal of Behavioral Decision Making*, 5, 169-186.

Figure caption

Kahneman, Ritov and Schkade (1999): Willingness to pay to restore damage to species that differ in popularity, as a function of the damage they have suffered. Schreiber and Kahneman (2000): Global evaluations of aversive sounds of different loudness as a function of duration, for subjects selected for their high sensitivity to duration. Novemsky and Kronzon (1998): Ratings of probability for predictions that differ in representativeness, as a function of base-rate frequency. Ariely (1998): Global evaluations of episode of painful pressure that differ in temporal profile, as a function of duration.


Kahneman, Ritov, & Schkade Data
35
30
25
20
15
10
5
0
0
20
40
60
80
Percentage Population Decline
Low
Medium
High
Schreiber & Kahneman Data
10
9
8
7
6
5
4
3
2
1
0
0
5
10
15
20
25
30
Duration in Seconds
71 dB
75 dB
78 dB
80 dB
Novemsky & Kronzon Data
Posterior Probability (%)
90
80
70
60
50
40
30
20
10
0
0
20
40
60
80
100
Base-rate (%)
Programmer
Surgeon
Accountant
Engineer
Ariely Data
80
75
70
65
60
55
50
45
40
35
30
0
10
20
30
40
50
Duration in Seconds
Down
Down&Up
Up
Up&Down